

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

NÁSTROJ PRO URČOVÁNÍ ATRIBUCE ONLINE MARKETINGOVÝCH AKTIVIT

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. IVAN ŠIŠÁK

BRNO 2013



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

NÁSTROJ PRO URČOVÁNÍ ATRIBUCE ONLINE MARKETINGOVÝCH AKTIVIT

TOOL FOR DETERMINING ATTRIBUTIONS OF ONLINE MARKETING ACTIVITIES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. IVAN ŠIŠÁK

VEDOUcí PRÁCE

SUPERVISOR

Mgr. KAMIL MALINKA, Ph.D.

BRNO 2013

Abstrakt

Nástroj umožňuje rozdělování zásluh na konverzi mezi různé marketingové aktivity. Jedná se o tzv. modelování atribuce. Rozdělení konverze je popsáno prostřednictvím modelů, kromě těch nejznámějších (model první interakce, model poslední interakce, lineární model), které umožňují vytvořit libovolné rozdělení konverze mezi marketingové aktivity. Použití tohoto nástroje může být vhodné při zjišťování efektivnosti marketingových aktivit.

Abstract

The tool allows distribution of credit for the conversion to various marketing activities. So-called attribution modelling is concerned. The distribution of conversion is described by the models, except those most famous (first interaction model, last interaction model, linear model), that allow to create any distribution of conversion to several marketing activities. The use of this tool may be useful in the surveys of the effectiveness of marketing activities.

Klíčová slova

webová analytika, atribuce, marketingová kampaň, modely atribuce, návštěva, návštěvník, zobrazení stránek, zdroje

Keywords

web analytics, attribution, attribution models, campaign, visit, visitor, page views, medium

Citace

Ivan Šišák: Nástroj pro určování atribuce online marketingových aktivit, diplomová práce, Brno, FIT VUT v Brně, 2013

Nástroj pro určování atribuce online marketingových aktivit

Prohlášení

Prohlašuji, že jsem tento semestrální projekt vypracoval samostatně pod vedením pana Mgr. Kamila Malinka, Ph.D. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....

Ivan Šišák
22. května 2013

Poděkování

Ďakujem za pomoc vedúcemu práce, ktorý mi pomáhal pri vypracovaní. Ďalej chcem poďakovať Mgr. Alexandre Gálovej, ktorá mi pomohla s korekciou textu.

© Ivan Šišák, 2013.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

1	Úvod	3
2	Špecifikácia projektu	5
2.1	Motivácia	5
2.2	Problémy	6
2.2.1	Určovanie unikátnosti návštevníka	6
2.2.2	Zakázaný Javascript	6
2.2.3	Doplňky v prehliadači	6
2.3	Metriky, dimenzie a segmentácia	7
2.4	Marketingové kampane	7
2.5	Zhrnutie riešených problémov	7
3	Webová analytika	8
3.1	Definícia webovej analytiky	8
3.2	Terminológia webovej analytiky	9
3.2.1	Stránka - Page	9
3.2.2	Volanie servera - Hit	9
3.2.3	Preklik - Click-Through	9
3.2.4	Zobrazenie stránky - Page view	10
3.2.5	Návšteva - Visit	10
3.2.6	Návštevníci alebo jedineční návštevníci	11
3.2.7	Nový návštevník - New visitor	11
3.2.8	Vracajúci sa návštevník - Return visitor	11
3.2.9	Opakovaný návštevník - Repeat visitor	12
3.2.10	Doporučiteľ - Referrer	12
3.2.11	Konverzia - Conversion	12
3.2.12	Konverzný pomer vs. pomer odchodov	13
3.3	Typy dát	13
3.4	Zdroje dát	15
3.4.1	Logovanie webového servera	15
3.4.2	Značkovanie stránok	17
4	Výpočet efektívnosti marketingových kampaní	20
4.1	Návrat investícií	20
4.2	Teória atribúcie	22
4.2.1	Modely atribúcie	22

5	Nástroje webovej analytiky a online marketingu	25
5.1	Nástroje webovej analytiky	25
5.2	Marketingové nástroje	27
5.2.1	PPC - platba za kliknutie	27
5.2.2	Pútač	28
5.2.3	Priamy marketing	29
5.2.4	Sociálne siete	29
5.2.5	Advergaming	29
6	Požiadavky a návrh aplikácie	30
6.1	Požiadavky	30
6.1.1	Funkčné požiadavky	30
6.1.2	Nefunkčné požiadavky	31
6.1.3	Prípady použitia	31
6.2	Komponenty aplikácie	31
6.2.1	Meranie dát	33
6.2.2	Uloženie dát	34
6.2.3	Analýza a zobrazenie dát	34
7	Implementácia	35
7.1	Implementácia zberu dát	35
7.1.1	Implementácia komunikácie medzi Elasticsearch a OWA	39
7.1.2	Databázový model	40
7.2	Implementácia získavania a analyzovania dát	44
7.2.1	Transformácia základných metrik	45
7.2.2	Transformácia s aplikovaním atribučného modelu	47
7.2.3	Problémy pri vývoji	49
7.3	Implementácia vytvárania štatistík	49
8	Testovanie a výstupy	50
8.1	Základné metriky	50
8.2	Konverzia s modelom	50
8.3	Konverzia bez modelu	52
9	Záver	54
9.1	Zhodnotenie projektu	54
9.2	Aktuálny stav	54
9.3	Budúcnosť projektu	55
A	Obsah DVD	58
B	Inštalácia	59

Kapitola 1

Úvod

Pod pojmom webová analytika si veľa ľudí predstavuje len štatistiky s počtom zobrazených stránok, prípadne s množstvom návštevníkov. V skutočnosti tento odbor zahŕňa oveľa viac. Pri použití webovej analytiky je dôležité nezostať len pri zbieraní dát o návštevníkoch, ale postupovať ďalej a dáta využiť vo svoj prospech, čo je hlavným cieľom použitia webovej analytiky podľa knihy [7]. Konkrétne napomáha získať poznatky, ktoré budú ďalej použité pri rozhodovaní akým smerom zamerať obchodný plán, do akých reklám investovať financie pre zvýšenie ziskov, akým spôsobom upraviť webové stránky, aby sa z návštevníkov stali zákazníci.

Aj keď je pomocou webovej analytiky možné získať rôzne poznatky, ako napríklad obľúbený obsah, množstvo a lojalita návštevníkov, geografická poloha návštevníkov, informácie o stránkach, z ktorých prichádzajú návštevníci, dosahovanie cieľov na webových stránkach, efektivita marketingových kampaní a iné, v rámci diplomového projektu bude pozornosť venovaná najmä efektívnosti marketingových kampaní, s ktorou úzko súvisí sledovanie cieľov, ktoré sú preddefinované webovou stránkou. Marketingové kampane slúžia na to, aby informovali o webovej stránke, o jej aktivitách, pričom ich hlavným cieľom je prilákať ľudí a mať z nich zisk. V tomto prípade nemusí ísť len o finančný zisk. Stačí si predstaviť stránku technickej podpory, kde užívateľ neodovzdá reálne peniaze, ale tým, že nájde riešenie svojho problému, dosiahne stránka cieľ, ktorý si definovala.

Pri určovaní efektivity marketingovej kampane rozhoduje, aký je pomer medzi investíciou do kampane a ziskom generovaným návštevníkmi, ktorí sa dostali na webovú stránku, vďaka tejto kampani a dosiahli cieľ, určený stránkou.

Kampaň sa podieľa na zisku. Cieľom diplomového projektu je určiť, ktorá kampaň a akou mierou sa na zisku podieľa. Pri jednakanálových kampaniach je výpočet jednoduchý, všetky zásluhy za zisk získa daná kampaň. Pri súbežných kampaniach je nutné zásluhy rozdeliť. Toto rozdelenie sa nazýva atribúcia a akým spôsobom sa delí podiel na zisku definuje pojem atribučný model. Viac sa tomuto problému venuje kapitola 4.

Toto je krátky a obecný pohľad na to, čomu sa diplomový projekt venuje. Detailnejšej charakteristike použitia a cieľov projektu sa venuje kapitola 2.

Kapitola 3 je venovaná popisu webovej analytiky. Presne vymedzuje daný pojem a predstavuje typy dát, ktoré možno získať a analyzovať. Ďalej pojednáva o zdrojoch, ktoré spomínané typy dát generujú a spôsoboch, ako sa tieto dáta získavajú. Keďže v oblasti webovej analytiky existuje niekoľko interpretácií jednotlivých pojmov, v tejto kapitole sú pojmy vymedzené tak, ako sa budú používať v celom texte práce.

Ako bolo spomenuté, cieľom projektu je zisťovanie efektívnosti marketingových kampaní. Ukazateľom efektívnosti marketingovej kampane je výpočet návratu investície, to

znamená koľkonásobne prevýšil zisk investíciu. Tento výpočet sa môže vykonávať ako odhad do budúcnosti pre zistenie veľkosti balíka financií nutného na dosiahnutie cieľa, no môže slúžiť aj ako pohľad do minulosti prípadne prítomnosti na určenie úspešnosti marketingovej kampane. Problému návratu investície a efektívnosti kampaní sa venuje kapitola 4.

Kapitola 5 popisuje konkrétne nástroje slúžiace na zber, analýzu a zobrazenie dát. Konkrétne sú popísané marketingové nástroje ako PPC - kampane preklikov, sociálna sieť, pútač a pod.

Kapitola 6 je venovaná funkčným a nefunkčným požiadavkom na aplikáciu. V tejto kapitole sa nachádza prepojenie jednotlivých komponentov, z ktorých sa aplikácia skladá.

Ďalšia kapitola potom popisuje implementáciu projektu. V nej sú predstavené možné riešenia zberu dát, z ktorých je jedno vybrané a použité v projekte. Okrem zberu je rozobratý dátový model, ktorý popisuje dáta ukladané v databázovom systéme. S ním je možné pracovať pri spracovaní dát.

Predposledná kapitola je venovaná testovaniu a predstaveniu jednotlivých výstupov aplikácie.

Kapitola 2

Špecifikácia projektu

V úvodnej kapitole je cieľ projektu popísaný všeobecne, táto kapitola sa mu venuje detailnejšie. Popisuje predpoklady, ktoré musia byť splnené, aby bolo možné cieľ dosiahnuť, diskutuje o problémoch, ktoré pri získaní dát nastávajú a navrhuje riešenia, ktoré tento projekt implementuje v praktickej časti. Kapitola taktiež popisuje požiadavky na metriky, typy marketingových kampaní a iné.

2.1 Motivácia

Každá webová stránka má istý obchodný plán, v ktorom je zahrnuté, čo chce návštevníkovi poskytnúť. Napríklad internetové obchody sa snažia predať tovar, spravodajské portály majú za cieľ informovať a portály technickej podpory majú riešiť problémy užívateľov. Na to aby sme mohli merať efektívnosť marketingovej kampane je nutné vedieť, aký je jej cieľ.

Ďalším predpokladom pre meranie efektívnosti kampane je konverzia. Užívateľ, ktorý príde na webové stránky, sa stáva návštevníkom. Pokiaľ splní preddefinovaný cieľ, tak sa z neho stane zákazník. Tejto premene z návštevníka na zákazníka sa hovorí konverzia.

Pri počítaní efektívnosti kampane sa vychádza z toho, koľko návštevníkov bolo po zhliadnutí prípadne po interakcii s istou reklamou konvertovaných na zákazníkov. Pri výpočte sa ráta s tým, že každá konverzia má svoju hodnotu. Napríklad pri nákupe tovaru to môže byť množstvo peňazí v rámci obchodnej transakcie, alebo môže byť hodnota konštantná, napríklad pri prečítaní článku alebo podaní istej informácie. Potom sa hovorí o tom, že vďaka danej kampani sa na webovej stránke vytvoril spomínaný zisk.

Ak stránka využíva jednu kampaň, priznanie hodnoty je jednoduché. Ak sa návštevník dostal na stránku pomocou danej kampane, každá hodnota konverzie je priradená tejto kampani. Pokiaľ však dôjde k situácii, že webová stránka využíva viacero kampaní, je nutné určiť, ku ktorej kampani bude táto hodnota priradená.

Modelová situácia by mohla vyzeráť nasledovne. Užívateľ zazrie prvú kampaň, pomocou ktorej príde na stránky, ale po istej dobe odíde, pričom nebol počas návštevy konvertovaný. O pár dní môže zazrieť inú kampaň, ktorá ho opäť dovedie na stránky, a dosiahne cieľ, čím návštevníka konvertuje. Teraz nastáva otázka, prečo nebol konvertovaný prvýkrát. Bola druhá kampaň skutočne lepšia, alebo len musel prvýkrát fyzicky odísť a tak nekonvertoval? Pokiaľ celú hodnotu konverzie dostane prvá kampaň, znamenalo by to, že žiadna iná nemala zásluhu na konverzii, čo platí aj opačne v prípade druhej kampane. Ak obe zohrali istú úlohu, mala by byť táto hodnota rozdelená. V podstate teória atribúcie vychádza zo zákona o akcii a reakcii a teda hovorí, že všetko má svoju príčinu. Takže by mala byť hodnota konverzie

rozdelená.

Delenie hodnoty konverzie je cieľom diplomového projektu. Nástroj na modelovanie atribúcie rozdelí z nameraných dát o návštevnosti a z analýzy dosiahnutých cieľov hodnotu konverzie podľa istých modelov a tým určí úspešnosť marketingových kampaní. Tento nástroj rozdelí hodnotu konverzie podľa modelov, ktoré sú popísané v kapitole 4.2.1.

2.2 Problémy

Keďže za určovaním atribúcie stojí meranie dát webovej analytiky, tak sú jej problémy aj problémami tohto projektu a je potrebné sa s nimi vysporiadať. V tejto kapitole sú zhrnuté len stručne, viac sa o nich možno dočítať v kapitole 3. V diplomovom projekte sú okrem týchto problémov popísané riešenia a ich implementácia.

2.2.1 Určovanie unikátnosti návštevníka

Určovanie unikátnosti je pri meraní dát vo webovej analytike jeden z najväznejších problémov. Určenie unikátnosti podľa IP adresy počítača nie je dostatočné, pretože nie každá pracovná stanica má svoju vlastnú IP adresu. Tento problém popisuje detailnejšie podkapitola 3.4.1. Druhým spôsobom je vygenerovať každému návštevníkovi jedinečný identifikátor, ktorý bude perzistentne uložený. Pre toto uloženie sa vo webovom prehliadači používajú cookies. Avšak ani tie nezaručujú jedinečnosť užívateľa, ale skôr jedinečnosť prehliadača, pomocou ktorého sú uložené.

Určenie jedinečnosti pomocou cookies je dostatočne presné pre použitie v projekte. Avšak v prípade, že sú cookies v prehliadači zakázané, môže dôjsť k nesprávnemu meraniu. Dnes väčšina užívateľov cookies nezakazuje, a preto to nie je pre tento projekt až tak závažný problém. Alternatívnymi uloženiami cookies by mohli byť napríklad ETag, ktorý je súčasťou protokolu HTTP a primárne slúži ako vyrovnávacia pamäť. Ďalšou metódou je tzv. lokálne úložisko, ktoré prichádza s HTML5. Dáta sú uložené na klientskej strane, ale nie sú prenášané v HTTP protokole, tak ako je to pri cookies.

2.2.2 Zakázaný Javascript

Merať dáta je možné pomocou Javascriptu alebo pomocou servera. Detailnejšie o týchto metódach hovorí sekcia ale aj kapitola o implemetácii projektu, kde sú predstavené rôzne riešenia a nakoniec je vybrané jedno, ktoré je implementované.

Problémom je, že pokiaľ sa využije pre meranie dát Javascript, nemuselo by dôjsť k žiadnemu meraniu a stráca sa akákoľvek informácia o užívateľovi. Preto je istejším spôsobom, ktorý sa v tomto projekte implementuje, meranie pomocou servera. Na webové stránky sa umiestni kód, ktorý na základe prijatého požiadavku zistí všetky informácie o užívateľovi a tieto dáta uloží do databázy. Týmto spôsobom sa projekt vyhne použitiu Javascriptu a riziku, že bude v prehliadači zakázané jeho spustenie.

2.2.3 Doplnky v prehliadači

Niektoré doplnky v prehliadači umožňujú blokovať odchádzajúce požiadavky, čím spôsobujú skreslené výsledky. Pokiaľ sa pri meraní využije serverové riešenie, takéto problémy nastanú, pretože server nepracuje s nastaveniami prehliadača.

2.3 Metriky, dimenzie a segmentácia

Pre projekt sú dôležité metriky, ktoré vyjadrujú plnenie cieľov, informácie o kampaniach a kanáloch, pomocou ktorých prichádzajú užívatelia na stránky. Metriky sú popísané v kapitole webovej analytiky, tu sú uvedené ako požiadavok, ktorý je nutné splniť, aby sa dala modelovať atribúcia. Ide o tieto metriky:

- odchodovosť pri plnení cieľa,
- konverzný pomer,
- hodnota konverzie daného cieľa,
- začatie daného cieľa,
- výnosy z obchodných transakcií,
- návštevy,
- zobrazenie stránky,
- konverzie.

Projekt by vo svojej vizualizačnej časti mal vedieť zobraziť tieto metriky vo vhodnej podobe, napríklad ako hodnotu, graf, tabuľku. Každá metrika je deliteľná dimenziou. Napríklad pri metrike konverzie môže dimenziu predstavovať zdroj návštevníkov, v tomto prípade to môže byť vyhľadávač, email a iné. Potom sa tento výsledok interpretuje ako počet konverzií rozdelených podľa zdroja. V projekte bude možné zobrazovať dáta podľa zvolenej podmienky, napríklad bude možné zobraziť konverzie, ktoré prišli z vyhľadávača Google a pod.

2.4 Marketingové kampane

Projekt sa pri skúmaní efektívnosti zaoberá webovými marketingovými nástrojmi. Neskúma kampane v novinách alebo televízii.

Pokiaľ boli v predchádzajúcich častiach textu nejasné pojmy, tak nasledujúce dve kapitoly definujú pojmy webovej analytiky a detailnejšie popisujú niektoré problémy spojené s meraním dát.

2.5 Zhrnutie riešených problémov

Nástroj pri implementácii bude riešiť nasledujúce spomenuté problémy:

- jedinečnosť návštevníka - identifikácia ukladaná pomocou cookies,
- merací kód na servery - k meraniu dát sa vykonáva na serveri a nie je nutné uvádzať o povolenom Javascripte, prípadne o problémoch s doplnkami,
- sledovanie online marketingových aktivít - neuvažuje sa o offline aktivitách akými sú noviny, televízia a pod.,
- metriky konverzie - počítané budú metriky spojené s kampaňami a konverziou, napríklad konverzia, hodnota konverzie, použité zdroje prevádzky a pod.

Kapitola 3

Webová analytika

Webová analytika ako odbor nie je definovaná štandardom a tak ani jednotlivé pojmy z jej oblasti nemajú pevne stanovený význam. Často dochádza k rozličným interpretáciám pojmov s ňou spojených. Preto je táto kapitola venovaná definíciám pojmov, aby nedochádzalo k ich zámene alebo k nesprávnej interpretácii. Pokiaľ sa v ktorejkoľvek časti textu použije niektorý z pojmov, jeho význam bude presne taký, ako je definovaný v tejto kapitole.

V úvode kapitoly je uvedená definícia pojmu webová analytika. Ďalej sú definované pojmy, ktoré sú v tejto oblasti dôležité a záver kapitoly je venovaný typom, zdrojom a metódam zberu dát.

3.1 Definícia webovej analytiky

Definícia pojmu webová analytika sa líši v závislosti od osoby, ktorá ju používa. Aj keď sa definície líšia, ich podstata je rovnaká. V nasledujúcom texte sú spomenuté tri definície.

Guy Creese, viceprezident Gartner Research, definuje pojem webovej analytiky [3] veľmi obecné:

„Sledovanie použitia webovej stránky a zobrazenie výsledkov sledovania tak, aby bolo možné lepšie pochopiť interakciu medzi návštevníkmi stránky a tým, čo stránka ponúka a využiť výsledky na zvýšenie lojality a predaja.“

Digital Analytics Association [18], známa aj pod názvom Web analytics association, definuje pojem nasledovne:

„Webová analytika je objektívne sledovanie, zbieranie, meranie, vyhodnocovanie a analyzovanie kvantitatívnych dát, ktoré optimalizujú webové stránky a webové marketingové aktivity.“

Definícia E.T. Petersona [11] rozširuje definíciu o typy dát, ktoré je možné zozbierať pri sledovaní a znie nasledovne:

„Webová analytika je zozbieranie rôznych dát zahŕňajúc dáta webovej prevádzky, webových transakcií, výkonnosti webového servera, štúdií použiteľnosti, užívateľsky odoslaných dát a ich súvisiacich zdrojov, ktoré majú pomôcť vytvoriť obecné pochopiteľný obraz o návštevníkovom správaní sa na webovej stránke.“

Každá citácia odpovedá na otázku ”čo je webová analytika” inak, ale majú aj niečo spoločné. Spoločným menovateľom je cieľ webovej analytiky, ktorým je z nazbieraných dát

získať čo najviac poznatkov, slúžiaci ako základ pre zvyšovanie kvality stránok po aplikácii optimalizačných metód. Keďže predmetom skúmania je čo najlepšie pochopiť, ktoré kampane ovplyvňujú a ako ovplyvňujú návštevníka, tak posledná definícia je tou, na ktorú tento text odkazuje. Pretože zahŕňa aj informácie z webovej prevádzky a z používania webových stránok, je na skúmanie správania najvhodnejšia.

Po zadefinovaní pojmu webovej analytiky je nutné spomenúť definície ďalších pojmov z tejto oblasti. Pri každom použití pojmu sa mieni jeho význam, tak ako je to popísané v nasledujúcej časti textu.

3.2 Terminológia webovej analytiky

Jednotlivé termíny webovej analytiky nie sú štandardizované a preto sa ich význam veľakrát odlišuje v závislosti od toho, aký nástroj sa pre webovú analytiku používa. Táto podkapitola definuje hlavné termíny, ktoré súvisia s webovou analytikou a pri ich použití vo zvyšných častiach textu sa odkazuje práve na ich význam definovaný v tejto kapitole.

V tejto časti sú spomenuté termíny od najnižšej úrovne (volanie servera, prekliky), ktoré majú nižšiu významovú hodnotu, cez metriky vyššej úrovne (zobrazenia stránok, stránky, návštevy) až po tú najvyššiu úroveň (konverzia).

Niektoré termíny sú uvedené v pôvodnom anglickom znení, pretože ich preklad by mohol do textu priniesť zmätok a pravdepodobne by nevystihoval ich podstatu. Termíny sú definované na základe knihy [11] a dokumentu [18].

3.2.1 Stránka - Page

Najzákladnejší pojem webovej analytiky je stránka. Jej definícia je nasledujúca:

„Stránka je dimenzia definovaná ako jednotka obsahu webovej aplikácie.“

Väčšinou sa stránkou myslí HTML dokument webovej aplikácie. Avšak niektoré nástroje webovej analytiky umožňujú definovať, ktoré požiadavky a typy súborov sú definované ako stránka. Takéto elementy sú potom zahrnuté v metrike zobrazenia stránok.

3.2.2 Volanie servera - Hit

Volanie servera je metrika, ktorá má pre analytika nízku hodnotu, lebo vyjadruje skôr technický aspekt webovej stránky. Volanie servera má nasledujúcu definíciu:

„Volanie servera je akcia na webovej stránke akou je zobrazenie stránky alebo stiahnutie súboru.“

V podstate ide o akúkoľvek požiadavku, ktorá je odoslaná na server. V prípade, že stránka obsahuje napríklad tri objekty, ktoré je nutné zobrazíť, tak to znamená, že sú vytvorené tri hity na objekty a jeden pre stránku (HTML dokument).

Hit sa využíva pri meraní záťaže servera, ale jeho použitie pri skúmaní správania užívateľa je nulové.

3.2.3 Preklik - Click-Through

Väčšinou sa táto metrika využíva pri určovaní efektívnosti interných alebo externých reklamných aktivít alebo pri zisťovaní správneho pochopenia navigácie na stránke.

„Metrika vyjadruje počet kliknutí na daný element stránky(odkaz, banner, kľúčové slovo).“

Často sa môže stať, že výsledky na strane, ktorá odosiela informáciu o kliknutí (reklamný server), sú iné ako výsledok, ktorý zmeral analytický nástroj. Malá odchýlka nevadí, ale veľké odchýlky by mali už podliehať hlbšiemu skúmaniu. Dôvodom odchýlok je to, že užívateľ omylom klikne na reklamu na stránke, ktorá ju zobrazila, ale na danú stránku sa nedostane, pretože preruší spracovávanie odoslanej požiadavky pomocou tlačidiel stop alebo späť v prehliadači. V tej chvíli mohol reklamný server kliknutie zaznamenať ale užívateľov príchod na stránku nebol zaznamenaný analytickým nástrojom stránky.

3.2.4 Zobrazenie stránky - Page view

Je to metrika, ktorá vyjadruje počet zobrazení stránky návštevníkom bez ohľadu na to, koľko obrázkov alebo objektov sa podieľa na vytváraní stránky.

Často býva problém rozpoznať, čo je stránka. Napríklad, ak aplikácia generuje stránky dynamicky, tak pohľad, ktorý zobrazuje produkty, môže byť tvorený šablónou. V takom prípade vzniká otázka, či pri zobrazení detailu sta rôznych produktov, kedy sa zobrazenie viaže na danú šablónu, má nástroj povedať, že bola stránka zobrazená stokrát, alebo má prehlásiť, že detail daného produktu bol zobrazený jeden krát. Pri sledovaní obľúbenosti produktu je lepšie, ak sa zobrazenie stránky počíta pre konkrétny produkt.

Ďalším problémom sú opakované požiadavky, napríklad pri resetovaní alebo obnovení stránky. Niektoré nástroje nezahrnú do výsledkov opätovné načítanie stránky, pokiaľ sa vykonalo pred vypršaním limitu (rádovo v sekundách) od pôvodného načítania stránky.

S ohľadom na všetky problémy spojené s identifikáciou stránky môžeme termín zobrazenia stránky definovať nasledovne:

„Zobrazenie stránky je počet úspešných nahratí dokumentu s obsahom, ktorý bol vyžiadaný užívateľom s ohľadom na spôsob doručenia a frekvencie vyžiadania.“

Termín, ktorý zoskupuje stránky do časového bloku, je návšteva.

3.2.5 Návšteva - Visit

Tiež je možné stretnúť sa s anglickým označením "session" alebo slovenským "sedenie". Definícia tejto metriky je nasledovná:

„Návšteva je interakcia návštevníka s webovou stránkou, pozostávajúca s nenulového počtu zobrazenia stránok. Ak návštevník nevykoná ďalšiu akciu na webovej stránke (napr. zobrazenie ďalšej stránky) do vypršania časového limitu, tak bude jeho návšteva ukončená.“

Typickým časovým limitom pre ukončenie návštevy je tridsať minút. Príklady počítania návštevy:

- Užívateľ kliká na stránke od 12:00, navštíví tridsať stránok a potom odíde. To je jedna návšteva.
- Užívateľ vloží do košíka produkty potom mu zazvoní telefón, telefonuje tridsaťjeden minút, potom sa vráti a dokončí nákup. Ide o dve návštevy, bol neaktívny príliš dlhú dobu, tým sa skončila jeho prvá návšteva a potom sa vytvorila nová.

- Užívateľ prechádza stránky náhodne, surfuje takto jeden a pol hodiny, potom prejde na inú stránku, ale stihne sa vrátiť do dvadsiatich deviatich minút. V takomto prípade ide stále o tú istú návštevu.

Návštevu vytvára osoba a pri spojení týchto dvoch termínov je nutné spomenúť návštevníka spolu s jeho rôznymi prívlastkami, ktoré vyjadrujú jeho vzťah ku stránke.

3.2.6 Návštevníci alebo jedineční návštevníci

Základným problémom pre určenie jedinečného návštevníka je určenie jeho jedinečnosti. V podkapitolách 3.4.1 je spomenuté, prečo by sa nemala IP adresa používať na určovanie jedinečnosti a tiež sa v podkapitole 3.4.2 prebrali problémy spojené s cookies, ale napriek tomu je ich použite presnejšie a preferovanejšie pre určenie jedinečnosti.

„Metrika jedinečných návštevníkov vyjadruje počet unikátnych osôb (nepočítajú sa vyhľadávacie roboty), ktorý za isté časové obdobie vykonali na webových stránkach aktivitu (vytvorili jednu alebo viac návštev).“

Každá osoba je do metriky v rámci obdobia započítaná len raz. Návštevníkov je možné rozlíšiť podľa vzťahu ku stránke na tri typy.

3.2.7 Nový návštevník - New visitor

Prvým typom návštevníka je nový návštevník.

„Metrika nových návštevníkov vyjadruje počet jedinečných návštevníkov, ktorí sú počas vybraného obdobia, pre ktoré sa výsledky počítajú, aktívni na stránkach a je to ich úplne prvá návšteva z celkového obdobia kedy sa dáta zbierajú.“

Každá osoba je v danom období označená za nového návštevníka iba raz a sčítaním nových a vracajúcich sa návštevníkov sa získa počet unikátnych návštevníkov. Táto metrika je vhodná pri zisťovaní, aká je lojalita jednotlivých návštevníkov.

Zaujímavé môže byť aj porovnanie interakcie so stránkou, ak sa dáta segmentujú na nových a vracajúcich sa návštevníkov. Analytikovi to ukáže, ako pracujú so stránkou užívatelia, ktorí s ňou majú skúsenosť a tí, ktorí sa na nej ocitli prvý krát.

3.2.8 Vracajúci sa návštevník - Return visitor

Vracajúci sa návštevník znamená za celé obdobie merania istú lojalitu voči webovej stránke a je to metrika, ktorá je vyjadrená nasledovným spôsobom:

„Metrika vracajúceho sa návštevníka vyjadruje v rámci vybraného obdobia počet unikátnych návštevníkov, ktorí v tomto období vykonali istú aktivitu a zároveň vykonali aktivitu aj v inom období, ktoré predchádzalo aktuálne sledovanému obdobiu.“

Opäť je každá osoba vzhľadom na sledované obdobie počítaná do tejto metriky len raz. Nie je možné, aby bola jedna osoba započítaná v rámci obdobia aj ako nový aj ako vracajúci sa návštevník.

3.2.9 Opakovaný návštevník - Repeat visitor

Posledným typom návštevníka je opakovaný návštevník. Metrika vyjadruje, aktivitu návštevníka v danom období. Jej presná definícia znie nasledovne:

„Metrika vyjadruje počet unikátnych návštevníkov, ktorý v rámci sledovného obdobia vykonali dve alebo viac návštev.“

Pri tejto metrike je možné, aby bol návštevník označený aj ako nový a zároveň ako opakovaný návštevník, alebo ako vracajúci sa a zároveň ako opakovaný návštevník.

3.2.10 Doporučiteľ - Referrer

Termín označuje zdroj, z ktorého sa návštevníci dostali na webovú stránku. Najčastejšími doporučiteľmi¹ sú URL alebo domény. Informácie o doporučiteľovi sú súčasťou hlavičky HTTP paketu. Problémom pri zisťovaní doporučiteľa je, ako zobraziť vo výsledkoch skutočnosť, v prípade, že táto informácia nebola uvedená v HTTP pakete. Väčšina nástrojov dopĺňa hodnotu „obľúbené“ alebo „záložka“, prípadne „priamo“ ak užívateľ zadal adresu ručne.

Definícia je pre tento termín nasledovná:

„Doporučiteľ je dimenzia, ktorá popisuje zdroj požiadaviek stránky alebo návštevy.“

Doporučiteľ môže byť nerozlíšená a úplná URL odkazujúca presne na stránku, z ktorej sa návštevník dostal na aktuálnu stránku. Ide o identifikáciu, ktorá hovorí o tom, z ktorých stránok sa návštevníci dostali na stránky po kliknutí na element v stránke (odkaz, banner, objekt).

Veľakrát sa stáva, že sú tieto URL spájané do skupín, ktoré vyjadrujú typ stránky. Skupiny môžu byť prispôbené v používanom nástroji. Najčastejšie sa používajú tieto skupiny:

- interný doporučiteľ - URL, ktorá je internou voči web stránke alebo skupina definovaná správcom analytického nástroja, napr. vo firmách môže byť niekoľko domén a používajú spoločný vyhľadávací nástroj,
- externý doporučiteľ - URL, ktorá je mimo web stránky,
- doporučiteľ je vyhľadávač - ide o doporučiteľa, ktorého URL bola vygenerovaná funkciou vyhľadávača,
- priama URL - tzv. žiadny doporučiteľ.

3.2.11 Konverzia - Conversion

Pre sledovanie výkonnosti marketingových aktivít je pojem konverzie jeden z najdôležitejších. Pre určenie konverzie je nutné definovať ciele. Tie sa definujú podľa toho, čo daná stránka poskytuje. V prípade internetového obchodu to je napríklad predaj produktu, v prípade portálu pre technickú podporu to je poskytnutie informácie, ktorú užívateľ hľadal atď.

¹v angličtine sa používa označenie referrer

Osoba, ktorá na webové stránky prišla, je návštevníkom, pokiaľ nevykonala aktivitu, ktorá by napĺňala jeden z definovaných cieľov. V prípade, že užívateľ vykoná danú aktivitu, stane sa zákazníkom. Táto premena z návštevníka na zákazníka je potom označovaná ako konverzia. Presná definícia je nasledovná:

„Konverzia je metrika, ktorá vyjadruje koľko krát bol úspešne splnený definovaný cieľ.“

Konverzia môže byť chápaná ako metrika v prípade, že sa hovorí o počte nákupov, registrácií a pod., ale taktiež môže byť dimenziou, keď chceme napríklad vyjadriť počet návštev spojených s istou konverziou.

S konverziou súvisia dva pomery, ktoré je vhodné v analytickom nástroji používať pre zistenie, koľko cieľov bolo začatých ale nedokončených.

3.2.12 Konverzný pomer vs. pomer odchodov

Nasledujúce pomery vyjadrujú úspešné a neúspešné dosiahnutie cieľa.

„Konverzný pomer je číslo vyjadrujúce počet, koľko krát bol daný cieľ dosiahnutý, vydelený vhodným deliteľom.“

Z definície je vyňatá informácia o deliteľovi a teda aj čitateľovi. Konverzný pomer môže byť počítaný s využitím počtu zobrazenia stránok, návštev alebo návštevníkov. Napríklad, ak je konverzný pomer viazaný na návštevy, tak vyjadruje počas koľkých návštev bol dosiahnutý cieľ vydelených celkovým počtom návštev.

S konverzným pomerom súvisí aj pomer odchodov, ktorý sa tiež môže nazvať pomer neúspešných konverzií.

„Pomer odchodov pre akýkoľvek krok v n-krokovom procese je jedna mínus počet jednotiek, ktoré sa objavajú v kroku n+1 podelených počtom jednotiek z kroku n.“

Opäť môže byť pomer počítaný nezávisle na použitých jednotkách, len musí platiť, že je jednotka v oboch krokoch rovnaká. Nemôžu sa použiť v n-tom kroku návštevníci a v n+1 kroku návštevy. Pri každom kroku je možné zistiť, aký bol úbytok, napríklad návštevníkov. Tým, že sa dosiahne v istom kroku vysoký úbytok, pravdepodobne je v ňom niečo, čo odrádza potenciálnych zákazníkov od ďalšieho používania webovej stránky.

3.3 Typy dát

Dát, ktoré by mohli pomôcť pri zvyšovaní kvality webovej stránky, je niekoľko. Podľa toho, ako sa získavajú, prípadne odkiaľ pochádzajú, je možné rozdeliť ich do piatich skupín [11]:

- **Dáta webovej prevádzky** - tradične je prevádzka zachytená v logovacích súboroch webového servera, ale v dnešnej dobe je tento zdroj už málo používaný a nahradil ho klientský prehliadač a kód v JavaScripte [7]. Kód je vložený do stránky a spustený pri jej otvorení, jeho hlavnou úlohou je zozbierať užitočné dáta a odoslať ich službe (nástroj pre webovú analytiku), ktorá s nimi ďalej pracuje. Najzaujímavejšie dáta tohto typu sú:

– počet zobrazení jednotlivých stránok,

- z akých stránok užívateľa prichádzajú,
 - kedy bolo najviac zobrazení stránok a iné.
- **Dáta transakcií** - sú typom dát, ktorý je odvodený z vykonaných transakcií na webovej stránke. V prípade internetového obchodu je pojem transakcie predstaviteľný najľahšie, preto uvedenie dát ako napríklad počet objednávok, predaj produktov, počet zákazníkov, zisk a ďalšie sú názornou inštancie dát tohto typu.
 - **Dáta výkonnosti webového servera** - sú dáta, ktoré nehovoria nič o priamej interakcii užívateľa so stránkou, ale poukazujú na to, akú záťaž dokáže zvládnuť webový server. Pokiaľ je odpoveď od serveru príliš pomalá, tak má užívateľ tendenciu stránku opustiť. Tým sa webové stránky pripravujú o potenciálny zisk. Arvind Jain v článku [8] hovorí o tom, že ľudia navštevujú stránky menej, ak nezaznamenajú žiadnu aktivitu stránky do 250 milisekúnd po odoslaní požiadavky. Pre porovnanie, žmurknutie trvá asi 400ms. Napríklad Jakob Nielsen [9] hovorí o tzv. 10-sekundovom pravidle:
 - desatina sekundy je čas, v ktorom by malo byť zrejmé, že systém odpovedá,
 - jedna sekunda je čas, v ktorom by mal užívateľ vidieť, že odpoveď systému je neprerušená,
 - desať sekúnd je čas, počas ktorého sa užívateľ sústreďí na vykonávanú úlohu. Potom má užívateľ vo zvyku úlohu prerušiť alebo je frustrovaný dlhou odpoveďou. To môže spôsobiť, že sa nikdy na stránky nevráti.
 - **Štúdie použiteľnosti** - je viac menej veda a jej pozorovanie sa zameriava na návštevníkov a aktuálnych zákazníkov a ich spôsoby pri používaní webovej stránky. Problémom je, že sa nedá presne určiť to, ako sa užívateľ cíti pri používaní danej webovej aplikácie, či je nový navigačný systém na stránkach lepší ako ten minulý. Avšak získané dáta môžu hovoriť o všeobecných princípoch, ktoré pozitívne vplývajú na užívateľa pri používaní webovej stránky.
 - **Užívateľský potvrdené dáta** - sú zozbierané priamo od užívateľa a presnejšie vyjadrujú jeho názor na konkrétnu vec. Získanie týchto dát riadi sám užívateľ, on sa rozhoduje, či dáta odošle a je na ňom, ako veľmi bude pravdovravný. Formy pri získaní týchto informácií sú rôzne a medzi najznámejšie patrí:
 - anketa - krátka otázka s výberom jednej alebo viac možností niekde za okrajom stránky,
 - dotazník - viacero otázok zameraných, buď na zistenie názoru na danú vec, alebo môže obsahovať otázky osobného charakteru (pre potreby segmentácie výsledkov) napr. vek, pohlavie,
 - komentár - dlhší textový popis návštevníka na danú vec, nie je možné automatizované spracovávanie.

Získanie informácie môže mať aj formu prieskumu alebo hlasovania, ktoré sa môžu zobraziť napríklad po nakúpení tovaru, alebo dokončení konkrétnej aktivity na webovej stránke. Tieto dáta majú oproti prvým trom typom tú výhodu, že skutočne odrážajú názor užívateľa na danú vec.

3.4 Zdroje dát

Na začiatku webovej analytiky, existoval jeden zdroj dát a teda bol automaticky považovaný za najlepší - išlo o logovacie súbory. O niekoľko rokov neskôr sa prišlo na to, že dáta z logovacích súborov neposkytujú všetky informácie, ktoré by bolo možné využiť pri zlepšovaní kvality webových stránok, preto muselo prísť nové riešenie. Novým zdrojom dát sa stali webové prehliadače, ktoré sú jednou stranou komunikačného kanálu medzi užívateľom a webovou stránkou alebo lepšie povedané podnikaním, ktoré poskytuje služby pomocou webovej stránky. Malý kus kódu JavaScriptu, ktorý je schopný pozbierať relevantné informácie a poslať ich do dátového centra, je začiatkom používania metódy "značkovanie stránky"².

Oba zdroje dát sú popísané v nasledujúcich dvoch častiach. Spomenuté sú ich výhody a nevýhody a podmienky kedy ich je možné použiť.

3.4.1 Logovanie webového servera

Použitie logovacích súborov nie je vlastnosť webového servera, ktorá je primárne určená pre použitie vo webovej analytike. Sú to textové súbory, do ktorých sú zapisované informácie o požiadavkách, ktoré musí webový server spracovať, aby vyhovel či už klientským staniciam, rôznym automatizovaným agentom napr. robotom vyhľadávača, ktoré indexujú obsah stránky, agentom pre monitorovanie výkonnosti alebo iným. Požiadavka nie je len zobrazenie webovej stránky v podobe štruktúrovaného dokumentu v jazyku HTML, ale ide aj o ostatné objekty, ktoré sa v HTML dokumente vyskytujú - obrázky, multimediálne súbory, skripty. V logovací súboroch je možné nájsť aj záznamy o chybách a mať prehľad o mŕtvych odkazoch na stránke, chybách vo webovej aplikácii, preťažení servera a pod.

Do logovacích súborov je možné zapísať pomerne veľké množstvo informácií, ale je nutné pamätať na to, že logovací súbor zaberá miesto na disku webového servera a čím viac návštevníkov prichádza, tým viac je nutné zapísať do webového logu. Preto je vhodné vybrať si také atribúty požiadavky, ktoré sú dôležité pri využití vo webovej analytike, aby sa zbytočne neplytvalo diskovým priestorom. Atribúty z logovacích súborov s významnou hodnotou sú:

- *požadovaný dokument* - dokument, ktorý chce užívateľ zobraziť vo webovom prehliadači,
- *dátum a čas* - je nutné určiť, kedy bola požiadavka vytvorená, aby sa mohli vytvárať sumáre za isté časové obdobie,
- *referencia* - informácia o tom, odkiaľ sa užívateľ dostal na danú stránku, dôležitá informácia pri zisťovaní vlastností príchodu na stránku,
- *HTTP stav* - chybový kód, dôležité pre získanie informácie o mŕtvych odkazoch,

Toto je výpis atribútov, ktoré majú v logovacích súboroch obecné najvyššiu hodnotu [11]. Existuje niekoľko atribútov, ktoré neboli spomenuté, ako napríklad IP adresa, klientská aplikácia, množstvo odoslaných bajtov, cookies. Je to z dôvodu, že ich použitie je závislé na zvážení istých alternatív. Napríklad IP adresa sa používa na určenie jedinečnosti klienta. To môže byť veľmi nepresné. Ako príklad možno uviesť ISP³, ktorý poskytuje internet tisíckam

²V angličtine sa táto metóda označuje ako "Page tagging".

³Poskytovateľ internetového pripojenia

domácností a stovkám podnikov, a tak je pravdepodobné, že tieto stanice sú pripojené pomocou technológie NAT⁴, ktorá mapuje pre zjednodušenie tisíce IP adries na jednu adresu a teda nie je možné konkrétne určiť z akého počítača požiadavka dorazila na webový server. Lepším riešením pre určovanie identity užívateľa je použitie cookies a uloženie jedinečného identifikátora ich pomocou, ale to vyžaduje dodatočné nastavenie na strane servera. Preto je nutné zistiť, či server podporuje cookies, ak áno, mala by byť pre určovanie jedinečnosti použitá práve táto technológia.

Pri množstve prenesených dát je nutné zodpovedať otázku, či je možné v použitom nástroji pre analýzu spracovávať informáciu o prenesených bajtoch, pokiaľ to možné je, tak sa môže táto informácia ukladať v logovacích súboroch a tým napríklad zistiť využitie šírky prenosového pásma serveru.

Aj keď logovacie súbory ako zdroj dát postupne miznú a sú nahradené značkováním stránok, majú stále svoje výhody a tie sú nasledovné:

- **Vlastníctvo dát** - pokiaľ firma prevádzkuje webové stránky na vlastných serveroch, tak má logovacie súbory vo svojom vlastníctve a spravujete si ich sama. Pokiaľ využíva hostingovú službu, môže sa stať, že nebude mať sprístupnené dáta a v tom prípade môže byť ich vlastníctvo skôr nevýhodou, pretože je to pre firmu cenný zdroj informácií o tom, ako sa jej v minulosti darilo.
- **Flexibilita dátového súboru** - keď je nutné získať informácie iba o žiadaných objektoch, čase a referenciách, tak sa do logovacích súborov nastaví zapisovanie len týchto informácií. Nie je potrebné zapisovať dáta, ktoré nie sú potrebné.
- **Jednoduchá implementácia** - väčšina, ak nie všetky webové servery, poskytujú možnosť zapisovať informácie o prevádzke do logovacích súborov.
- **Možnosť logovať rovno do databázového úložiska** - niektoré otázky o návštevníkoch môžu byť zodpovedané pomocou SQL výrazu.
- **Schopnosť zistiť úplné stiahnutie objektu** - možno zistiť, či bol súbor stiahnutý celý alebo užívateľ prerušil sťahovanie, alebo sa vrátil na predchádzajúcu stránku.
- **Možnosť merať dátový tok od agentov** - keďže cieľom optimalizácie pre vyhľadávacie (používa sa skratka SEO⁵) je zaujať čo najviac vyhľadávacích robotov, táto informácia môže byť vhodná pre zistenie efektivity SEO.

Po uvedení niekoľkých výhod je vhodné spomenúť aj nevýhody logovacích súborov:

- **Vyrovňacia pamäť prehliadača** - prehliadače si ukladajú nedávno zobrazený obsah, čiže keď užívateľ stlačí vo svojom prehliadači tlačidlo späť alebo ďalej, tak je tento obsah obslužený z vyrovnávacej pamäte. Je to veľmi prívetivé pre užívateľa, ale pri získaní dát spôsobuje nepresnosti. Pri meraní počtu zobrazení stránky chceme vedieť koľko krát sa zobrazila, avšak keď dôjde na strane užívateľa k nahratiu stránky z vyrovnávacej pamäti, tak sa informácia o zobrazení stránky stráca, pretože v tej chvíli prehliadač nekomunikuje so serverom. Riešením je Cache-busting [12], ktorý zabráňuje prehliadaču obslužiť stránku z vyrovnávacej pamäte.

⁴Preklad sieťových adries

⁵ SEO je skratka zo slov Search engine optimization

- **IP adresa ako unikátny identifikátor** - už bolo spomenuté, prečo IP adresa nie je spoľahlivý identifikátor jedinečnosti, ale navyše môže IP adresa viesť aj k nejednoznačnosti pri určovaní oblasti, z ktorej prichádzajú požiadavky. Veľký ISP má svoje proxy servery rozmiestnené po svete a ak je užívateľ fyzicky v Českej republike, môže byť požiadavka na serveri vyhodnotená ako pochádzajúca z úplne iného regiónu.
- **Roboty vyhľadávačov** - pri zisťovaní návštevnosti môže byť veľmi mätúcim faktom, že sa do štatistiky počítajú aj požiadavky od automatizovaných agentov, ako sú napríklad roboty vyhľadávačov, ktoré indexujú obsah stránok. Predsa len, keď chce majiteľ webovej aplikácie vedieť návštevnosť, tak ho nezaujíma množstvo požiadaviek od robotov ale od ľudí. Preto je nutné tieto informácie filtrovať z logovacích súborov pri analýze, alebo ak to server umožňuje, tak ich vôbec nezapisovať.

3.4.2 Značkovanie stránok

Keď sa hovorí o zdrojoch dát pre webovú analytiku, značkovanie stránok dnes nahradzuje logovacie súbory [7]. Dôvod je najmä ten, že tento spôsob umožňuje získať oveľa rozmanitejšie dáta, napríklad o operačnom systéme, prehliadači, o aktivite užívateľa (akým spôsobom sa pomocou kurzora myši pohyboval na stránke, o elementoch dokumentu, na ktoré klikol atď).

Princíp ako funguje značkovanie stránok je možné zhrnúť do nasledujúcich krokov [11]:

- Nástroj pre značkovanie poskytuje v jazyku JavaScript kód, ktorý sa vloží do sledovaných stránok. Často je tento kód konfigurovateľný a je možné nastaviť niektoré premenné, ktoré nástroj podporuje napr. priradenie typu stránky podľa kontextu.
- Návštevník si v prehliadači otvorí stránku, kde sa takýto kód nachádza.
- Keď je stránka stiahnutá, tak sa spustí kód vložený v stránke, ktorý zoskupí dáta do jednej požiadavky, ktorú odošle na server. Veľakrát je požiadavka odoslaná ako žiadosť o obrázok, ktorý má rozmer 1x1 pixel. URL obrázku je doplnená o reťazec⁶ obsahujúci dáta, ktoré sa majú zaznamenať.
- Analytický nástroj prijme požiadavku na takýto obrázok, rozkóduje informáciu z URL a uloží tieto dáta do databázového systému.
- Obrázok je odoslaný späť klientovi a tým je celá transakcia ukončená.
- Klient obrázok prijme, ale ten sa nezobrazí nikde na webovej stránke.

Výhody značkovania stránok je možné zhrnúť do nasledujúcich bodov [11].

- **Presnosť** - keďže sa dáta získavajú na strane klienta a nie na strane servera, je možné doceliť vyššiu presnosť dát v porovnaní s logovacími súbormi. Väčšina riešení pre značkovanie používa pre identifikáciu užívateľa namiesto IP adresy cookies. Taktiež využívajú klientské stanice Cache-busting [12], čím sa zabezpečí, že sa dáta zaznamenajú aj pri stlačení tlačidla späť a ďalej v prehliadači.
- **Návštevy automatizovaných agentov** - väčšina týchto agentov nespúšťa JavaScriptový kód, a tak nie je ich návšteva stránky odoslaná do aplikácie na zber dát.

⁶V angličtine sa označuje ako query string

- **Rýchlosť výsledkov** - prijaté dáta sú okamžite rozkódované a uložené do databázy, teda nie je nutné dodatočné spracovávanie. Dáta sú okamžite pripravené na zobrazenie výsledkov.
- **Flexibilita dátového súboru: premenné** - okrem základných dát je možné definovať aj ďalšie premenné zobrazovanej stránky, napr. názov stránky, skupina podľa obsahu stránky, ID kampane, zaradenie návštevníka, produkt, kategória produktu, značka, cena atď.

Aj keď sa zdá, že značkovanie stránok je dobrý spôsob ako získať väčšie množstvo informácií, s vyššou hodnotou a vyššou presnosťou tak existuje niekoľko nevýhod, ktoré môžu byť veľmi významné pri rozhodovaní, či značkovanie použiť alebo nie [11].

- **Závislosť na JavaScripte a Cookies** - pri značkovaní sa väčšina nástrojov spolieha na použitie JavaScriptu, ktorý má za úlohu získať čo najviac informácií o návštevníkovi, a cookies, ktoré slúžia ako úložisko pre udržanie, napríklad identity užívateľa, identifikácie návštevy a ďalších. Ak sú tieto technológie zakázané, môže byť získanie informácií nemožné alebo čiastočne obmedzené, čo môže spôsobiť nepresnosti v dátach. Keďže nie je možné spustiť kód v JavaScripte, tak nie je možné ani získať informácie o užívateľovi a keď nie sú povolené cookies, môže byť identita užívateľa stále iná a pri 50-tich zobrazeniach stránky tým istým návštevníkom sa získa informácia, že ide o 50 nových, jedinečných návštevníkov. Podľa výskumu [20] má len veľmi málo ľudí zakázaný Javascript vo svojich prehliadačoch a teda chyba, ktorá sa môže vyskytnúť vo výsledných dátach je zanedbateľná.
- **Cookies ako identifikátor jedinečnosti** - veľa nástrojov spomínajúc návštevníkov má na mysli ľudské bytosti, ale v skutočnosti je tomu úplne inak. Pri zbieraní dát sa zbierajú dáta o webovom prehliadači. Dnes je bežné, že sa užívateľ pripája na webové stránky z niekoľkých zariadení a nie je jednoduché aby bola jeho identita (vzhľadom na užívateľa je to nemožné) na všetkých zariadeniach rovnaká. Každý prehliadač má svoj vlastný súbor cookies a ak sa užívateľ pripojí na viacerých zariadeniach, tak je zakaždým jeho identita iná. Na druhú stranu pokiaľ použije zariadenie po tom, čo ho používala iná osoba, preberá identitu tohto návštevníka. Nástroje predpokladajú vzťah 1:1 medzi človekom a webovým prehliadačom, ale tento vzťah je v skutočnosti N:N. Dôvody, ktoré môžu ilúziu vzťahu 1:1 narušovať sú nasledujúce:
 - užívatelia používajú viacero zariadení = viacero prehliadačov,
 - prehliadač na danom počítači nepoužíva jedna konkrétna osoba,
 - užívatelia zakážu použitie cookies,
 - ľudia zmažú vyrovnávaciu pamäť pre cookies,
 - ľudia aktualizujú svoj počítač, napr. preinštalovanie bez zálohy dát.
- **Pridanie značkovania do každej stránky** - pri statických webových stránkach alebo aj pri zle vytvorených dynamických stránkach môže byť veľká nevýhoda. Pri statických stránkach musia byť značky vložené do každej stránky ručne a pokiaľ ich je niekoľko stoviek, je proces nastavovania merania veľmi zdĺhavý.
- **Výkon** - čokoľvek sa pridáva do webovej stránky, prináša svoju cenu v podobe množstva prenesených dát a v podobe využitia zdrojov pri spustení kódu pre získavanie dát. Aj keď pri dnešných prenosových rýchlostiach netrvá stiahnutie súboru dlho, zbytočné plýtvanie zdrojmi servera nie je žiaduce.

- **Typy dát** - logovacie súbory môžu veľmi jednoducho sledovať aj sťahovanie rôznych súborov (PDF, obrázky, komprimované súbory), pri značkovaní je toto možné dosiahnuť tiež, ale je nutné, aby nástroj podporoval sledovanie kliknutí na stránke, prípadne aby sledoval aj iné požiadavky, ktoré boli vytvorené pri zobrazení aktuálnej stránky.

Táto kapitola položila terminologický základ pre ďalšiu prácu. Po zedefinovaní termínov z oblasti webovej analytiky bude v nasledujúcej kapitole predstavené určovanie efektívnosti marketingových kampaní, atribúcia a atribučné modely.

Kapitola 4

Výpočet efektívnosti marketingových kampaní

Prechádzajúca kapitola vysvetľuje niekoľko pojmov, ktoré je nutné poznať, keď sa chce človek orientovať v oblasti webovej analytiky. Slúži na vysvetlenie pojmov a k vysvetleniu problémov, ku ktorým pri zbere dát dochádza.

Cieľom projektu je určovanie atribúcie marketingových aktivít. To, čo vlastne atribúcia je, o tom hovorí táto kapitola. Okrem vysvetlenia pojmu atribúcie sú vysvetlené modely, ktoré sa pri nej využívajú.

Atribúcia sa berie ako nástroj na rozdeľovanie zásluh na konverzii medzi jednotlivé kanály, pomocou ktorých sa návštevník dostáva do styku s webovou stránkou. Výpočet návratu investície poskytuje vyjadrenie toho, aký bol pomer medzi investíciami do marketingovej kampane a výnosmi, ktoré kampaň priniesla. Týmto výpočtom sa presne ekonomicky vyjadří úspešnosť kampane. Aj keď je návrat investície kľúčový pri určovaní efektivity marketingovej kampane, v aplikácií sa jeho výpočet neobjaví. No vzhľadom na jeho dôležitosť je v tomto texte spomenutý.

Táto kapitola opisuje obidva spomenuté problémy – *návrat investícií*¹ a *atribúcie*.

4.1 Návrat investícií

Definícií existuje niekoľko, ale v projekte sa uspokojíme s tou úplne základnou. Podľa [17] ROI je podiel medzi výnosmi a množstvom investícií. Vynásobením získaného čísla číslom 100 dostaneme výsledok o koľko percent je vyšší návrat oproti investíciám. Vzorec 4.1 popisuje spomínaný výpočet.

$$ROI = \frac{vynos}{investicia} \times 100 \quad (4.1)$$

Návrat investície môže byť počítaný ako:

- **predpoveď** - na základe odhadovaných dát, akými sú konverzný pomer, priemerná hodnota objednávky, či očakávaný výnos sa odhadne veľkosť investície do kampane,
- **vyhodnotenie** - na základe nameraných dát sa zistí, aké boli výnosy a ktoré kampane sú efektívne a ktoré nie.

¹v angličtine často používaná skratka ROI - Return of investments

Postup počítania návratu investície je možné zapísať do niekoľkých krokov. V prípade odhadovania by sa postupovalo nasledovne [11]:

- Stanovenie celkovej hodnoty, ktorú má kampaň dosiahnuť.
- Zistiť priemernú hodnotu objednávky alebo priemernú hodnotu konverzie.
- Stanovenie požadovaného násobku investície.
- Vydeliť celkový výnos z kampane koeficientom stanoveným v predchádzajúcom bode pre stanovenie maximálnej možnej investície do kampane.
- Stanovenie konverzného pomeru kampane na základe priemerného konverzného pomeru alebo na základe predchádzajúcich podobných kampaní.
- Vydeliť cieľový výnos kampane pomocou priemernej hodnoty objednávky alebo konverzie, aby sa získalo množstvo konverzií potrebných na získanie cieľových výnosov z kampane.
- Vynásobiť počet konverzií konverzným pomerom na získanie potrebného množstva respondentov.
- Vydeliť celkovú investíciu do kampane počtom respondentov potrebných na kampaň, aby sa získala veľkosť investície v prepočte na jedného respondenta.

Pre lepšie predstavenie je uvedený nasledujúci príklad.

- Manažér chce zarobiť 10 000 korún.
- Priemerná cena objednávky je 100 korún.
- Investícia do kampane sa má vrátiť 5 krát.
- Do kampane nechce investovať viac ako 2 000 korún.
- Konverzný pomer je 5%.
- Potrebných je 100 konverzií, aby zarobil 10 000 korún.
- S konverzným pomerom 5% je potrebných 2 000 respondentov.
- S maximálnou investíciou do kampane 2 000 korún a s potrebnými 2 000 respondentmi na dosiahnutie cieľa nemôže zaplatiť manažér za jedného respondenta viac ako jednu korunu.

Takýto odhad môže poskytnúť nielen informáciu o tom, či sa oplatí investovať do zvolenej kampane, ale taktiež môže po jej skončení poskytnúť informáciu o tom ako úspešne boli odhadnuté investície a využiť ju pri plánovaní ďalších kampaní.

V prípade zisťovania efektívnosti prebiehajúcej kampane sa návrat investície počíta z nameraných hodnôt. K výpočtu stačí zistiť len 3 metriky a tými sú celková cena kampane, počet respondentov a počet konverzií. K tomu je ešte nutné stanoviť priemernú hodnotu, ktorá sa získa pri jednej konverzii.

- Zistiť celkovú investovanú cenu do kampane.

- Zistiť počet respondentov, ktorý na kampaň reagovali.
- Zistiť počet konverzií, ktoré boli vykonané v rámci tejto kampane.
- Stanovenie priemernej hodnoty jednej konverzie.
- Vynásobením počtu konverzií a priemernej hodnoty jednej konverzie sa získa celkový výnos.
- Odčítaním celkového výnosu a celkovej investovanej ceny do kampane sa získa celkový zisk.
- Podielom celkového zisku a investície sa získa efektívnosť kampane.

Týmto je načrtnutý úvod do výpočtu efektívnosti kampane. Nespomína sa nič o tom, či bola konverzia rozdelená medzi všetky kanály, ktoré sa na nej podieľali. Pravdepodobne bola konverzia priznaná vždy len jednému kanálu. Avšak veľa webových stránok používa niekoľko kanálov na upútanie návštevníkov s cieľom dosiahnuť konverziu. V takom prípade je vhodné rozdeliť konverziu medzi viacero kanálov. Dôvodom toho je teória atribúcie, ktorá je popísaná v nasledujúcej podkapitole.

4.2 Teória atribúcie

Dôvodom rozdelenia konverzie medzi viacero kanálov pre efektívnejšie zisťovanie ich výkonnosti je vlastne myšlienka, na ktorej je teória založená. Táto teória je založená na myšlienke, že každá akcia má svoju reakciu [6].

Teória atribúcie vyjadruje, že ľudia sa pokúšajú zistiť prečo druhí ľudia robia to, čo robia [5]. Atribúcia je sociálno-psychologický pojem, ktorý vysvetľuje, že ľudia sa pozerajú na svet a nevnímajú ho ako úplne náhodný ale snažia sa pripisovať veciam istý význam.

Pri aplikácii predchádzajúcich viet na webovú analytiku je v podstate tou „vecou“ vo svete práve konverzia, pri ktorej následne dochádza k otázkam, prečo k nej došlo. Príčinami konverzie sú kanály, ktorými si návštevník prešiel, pokiaľ dosiahol konverziu, pretože práve kanály ho ovplyvnili, upútali.

4.2.1 Modely atribúcie

Atribučný model [16] je pravidlo alebo súbor pravidiel, ktoré určujú, ako bude medzi kontaktné body na konverzných trasách rozdelený kredit za predaj a konverziu.

Konverznou trasou sa rozumie sekvencia kanálov (kontaktných bodov), ktoré návštevník použil počas cesty, na ktorej nakoniec splnil cieľ stránky a vytvoril konverziu.

Modelová situácia by mohla byť taká, že užívateľ prišiel na webové stránky pomocou sociálnej siete, kde videl príspevok. V tomto prípade je prvým kanálom sociálne médium. Po troch dňoch začal zisťovať pomocou vyhľadávača ceny produktu na iných stránkach a porovnával ich so sledovanou stránkou. Druhým kanálom na ceste ku konverzii je vyhľadávač. Na ďalší deň príde zákazník na stránky po tom, čo klikol na reklamu v emaily, ktorá je súčasťou kampane a nakúpi v obchode na stránkach - dôjde ku konverzii. Na tejto situácii budú vysvetlené modely, ktoré rozdelia konverziu podľa definovaných pravidiel.

Model poslednej interakcie

Tento model priradzuje všetky zásluhy (kredit) poslednej kampani (email), pomocou ktorej návštevník prišiel na stránky a bol konvertovaný. Podľa výskumu [16] je tento model veľmi neefektívny, pretože priradzuje váhu len jednej kampani, tej poslednej. Na druhú stranu môže byť výhodný pre určenie, ktorá kampaň najviac vplývala na návštevníka, keď došlo konečne ku konverzii.



Obrázek 4.1: Model poslednej interakcie

Model prvej interakcie

Model priradzuje všetku váhu prvej kampani (sociálne médium), pomocou ktorej prišiel na stránky a bol konvertovaný. Existuje tu podobný problém ako pri prvom modeli [16] teda zásluhy nie sú rozdelené medzi ďalšie kampane. Avšak tento model je vhodný pre zistenie, ktorá kampaň pomáha pri objavovaní stránok.



Obrázek 4.2: Model prvej interakcie

Lineárny model

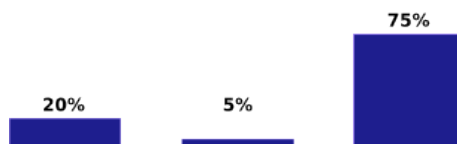
Zásluhy sú rovnomerne rozdelené medzi jednotlivé kampane, ktoré návštevníka pritiahli, až pokiaľ nevykonan konverziu [16]. V podstate to znamená, že každá kampaň má na návštevníka rovnaký vplyv, bez ohľadu na to kde v ceste za konverziou bola vykonaná interakcia medzi kampaňou a zákazníkom.



Obrázek 4.3: Lineárny model

Pozičný model

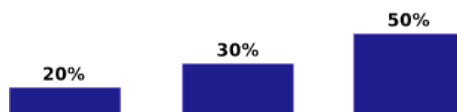
Pozičný model priradzuje váhu podľa pozície kampane v konverznej ceste. Napríklad pri rozdelení zásluh 20% : 5% : 75% dostane sociálne médium 20% kreditu, vyhľadávač 5% a email 75%.



Obrázek 4.4: Pozičný model

Časovo meniaci sa model

Myšlienkou tohto modelu je priradiť vyšší kredit kanálom, ktoré sú bližšie ku konverzii. Tento model sa využíva pri krátko trvajúcich kampaniach, kedy je nutné zistiť, či boli tieto kampane úspešné a potláča kampane, ktoré boli použité dávnejšie.



Obrázek 4.5: Časovo meniaci sa model

Týmto boli predstavené základné atribučné modely. V implementácii bude vytvorený model, ktorý bude možné nastaviť tak, aby sa zásluhy mohli rozdeliť podľa každého zo spomenutých modelov.

Po tejto zmienke o nástroji zostáva popísať nástroje webovej analytiky a rôzne marketingové nástroje, ktoré sú aktuálne dostupné. Nasledujúca kapitola sa venuje práve týmto dvom témam.

Kapitola 5

Nástroje webovej analytiky a online marketingu

Táto kapitola sa zaoberá dvomi skupinami nástrojov, ktoré je možné vo webovej analytike používať. Prvá je skupina nástrojov, pomocou ktorých je možné zozbierať rôzne typy dát o návštevníkoch, vyťažnosti stránok, kampaniach, konverziách atď. Okrem zozbierania dát ponúkajú tieto nástroje aj ich analýzu dát a grafické zobrazenie pomocou grafického užívateľského rozhrania. V spomenutých prípadoch ide o webové užívateľské rozhranie.

Druhú skupinu tvoria marketingové nástroje, ktoré majú za úlohu upútať návštevníka a za účelom konverzie ho priviesť na stránky.

5.1 Nástroje webovej analytiky

Sú to nástroje s trojitou funkcionalitou:

- meranie - získavanie informácií pomocou prehliadača a značkovania stránok,
- analýza - analyzovanie dát a vytváranie štatistických súborov,
- zobrazenie výsledkov - pomocou grafického rozhrania prístup k analyzovaným dátam.

Existuje veľké množstvo nástrojov. Ich čiastočný zoznam je možno vidieť na stránkach www.aboutwebanalytics.com. Líšia sa najmä funkcionalitou, uložením dát a cenou. Táto práca sa venuje nástrojom, ktoré sú podľa spomínaných stránok ohodnotené ako nástroje s nízkou cenou a s celkovým dobrým hodnotením. Ide o nástroje Google Analytics, Piwik či Open web analytics.

V tabuľke 5.1 je prehľad funkcií, ktoré jednotlivé nástroje ponúkajú. Sú to komplexné a funkcionalitou veľmi podobné nástroje. Podľa porovnania je na tom najlepšie nástroj Google analytics, ktorý ponúka veľa možností, či už v rámci sledovania alebo analýzy. Piwik je veľmi porovnateľný nástroj s Open web analytics, ale užívateľské rozhranie je pomalé a niektoré metriky je nutné doimplementovať. Open web analytics má prívetivejšie a intuitívnejšie rozhranie, aj keď to je vecou subjektívneho názoru.

Najväčšiu hodnotu pri podnikaní majú užívateľské dáta [7]. Preto je veľmi dôležité, aby mal správca analytického nástroja pod kontrolou zozbierané užívateľské dáta. Google analytics neumožňuje prístup k suchým dátam a preto je aj ďalšia práca s nimi náročná.

Open web analytics umožňuje zbierať dáta aj o pohyboch užívateľa na danej stránke tzv. "heatmaps". Dáta je možné sledovať nie len na strane klienta pomocou kódu v jazyku

Vlastnosti	Google Analytics	Piwik	Open web analytics
Sledovanie			
Vyhľadavanie	•	•	•
Spracovanie v reálnom čase	•	•	•
E-mailové kampane	•	•	•
Sociálne médiá	•	•	•
Mobilné OS	•	•	×
Sledovanie viacerých domén	•	•	•
Udalosti	•	•	•
E-commerce	•	•	•
Užívateľom definované premenné	•	•	•
Cena kliknutí (importovanie)	•	•	•
Unikátni návštevníci	•	•	•
Prispôsobené prekliknutie sa	•	•	•
Analýza			
Vizualizácia procesov	•	×	•
Rozdelenie kreditu medzi viacero kampaní	•	•	•
Lojalita návštevníkov	•	•	•
Správanie na stránke	•	•	•
Návštevníkova cesta	•	•	•
Segmentovanie	•	•	•
Filtre	•	•	•
Predikatívnosť	•	×	×
Interné vyhľadavanie	•	×	•
Sekundárne dimenzie	•	•	×
Prispôsobenie štatistík	•	•	•
Export			
PDF a Email	•	•	•
Excel	•	•	•
CSV, XML	•	•	•
Databázové dáta	×	•	•
API	•	•	•

Tabulka 5.1: Prehľad vlastností nástrojov webovej analytiky

Javascript, ale taktiež na strane servera pomocou kódu v jazyku PHP . V porovnaní s Google analytics umožňuje Piwik aj Open web analytics sledovať neobmedzené množstvo stránok [10]. Google analytics umožňuje sledovať 50 stránok na jeden užívateľský účet.

Pre tento projekt bude najvhodnejší nástroj Open web analytics, ktorý poskytuje v porovnaní s ostatnými nástrojmi viacero možností sledovania, je to projekt s otvorenou licenciou a hlavne umožňuje prácu so suchými dátami.

Toto krátke porovnanie ukazuje rozdiely medzi jednotlivými nástrojmi a objasňuje dôvody použitia Open web analytics v tomto projekte. Len pre zhrnutie, dôležitým faktom, ktorý zohráva najväčšiu úlohu a vylučuje použitie Google analytics, je, že Open web analytics umožňuje prácu so suchými dátami. Nie je nutné pracovať s dátami len na úrovni API a vyžiadať si dáta len v podobe preddefinovaných metrík a dimenzií, ale je možné pracovať s dátami priamo, tak ako sú uložené v databáze.

V nasledujúcej časti sú popísané marketingové nástroje, ktoré sa využívajú na upútanie potenciálneho zákazníka a dávajú vedieť o webovej stránke, napríklad na reklamných miestach, vo vyhľadávačoch alebo na weboch zobrazujúcich reklamu.

5.2 Marketingové nástroje

Marketingové nástroje sa rozdeľujú na online a offline. Tie s označením offline môžu byť reklamy v printových médiách, televízii, billboardy a pod.

Online marketingové kampane sú napríklad PPC alebo bannery. Tieto sú popísané v nasledujúcej časti textu. Tieto reklamy majú oproti offline reklamám jednu veľkú výhodu, sú merateľné. Je možné zaznamenávať koľko ľudí využilo pri prístupe na webové stránky danú reklamu a tým zistiť, či sa vynaložené prostriedky vracajú.

Marketingové nástroje sú spojené s pojmom SEM [15], ktorý znamená marketing vo vyhľadávačoch. Je to komplexný marketingový nástroj zameraný na propagáciu a zvyšovanie viditeľnosti webu. Na rozdiel od SEO, SEM nevyžaduje žiadne zásahy do webu a teda je možné využívať ho behom pár minút. Tento nástroj umožňuje osloviť zákazníkov až vo chvíli keď daný produkt, informáciu alebo službu skutočne potrebujú.

Jednou formou marketingu vo vyhľadávačoch je aj PPC nástroj, pri ktorom zadávateľ reklamy platí za reklamu podľa toho, koľko krát bolo na reklamu kliknuté. Ďalšími formami, ktoré súvisia menej alebo vôbec so SEM, sú pútače, priamy marketing, sociálne média, advergaming, microsites a iné. Jednotlivé nástroje sú popísané v nasledujúcich častiach textu.

5.2.1 PPC - platba za kliknutie

Nástrojom spadajúcim pod SEM výraz je PPC reklama. PPC¹ kampane sú marketingovým nástrojom, pri ktorom inzerent neplatí za umiestnenie alebo zobrazenie reklamy, ale platí až vo chvíli, keď užívateľ klikne na daný odkaz.

Inzerent si založí účet u poskytovateľa tohto druhu reklamy a pomocou administratívneho rozhrania je možné pridávať nové a spravovať existujúce reklamy. Potom inzerent nastaví, koľko je ochotný za reklamu zaplatiť. Čím viac je ochotný zaplatiť, tým vyššie je v odkazoch vo vyhľadávači a tým častejšie sa reklama zobrazuje. Výhodami PPC kampaní sú [21]:

- je presne cielená,

¹Skratka z anglických slov Pay per click

- neobťažuje, pokiaľ si ju užívateľ nechce všímať, tak sa mu nijak nevnučuje,
- je platená za prístup na inzerovanú stránku, nie za zobrazenie reklamy,
- prínos kampane je možné merať.

PPC - SKlik

SKlik je nástroj pre PPC kampane od Seznam.cz. Inzercia, ktorá je poskytovaná pomocou SKlik, je zobrazovaná na dvoch sieťach. Prvou sieťou je sieť vyhľadávania, teda stránka s výsledkami vyhľadávania na seznam.cz, alebo v ďalších službách tejto spoločnosti, ktorými sú Zbozi.cz, Firmy.cz, Sbazar.cz atď. Druhou sieťou, kde sa reklama vyskytuje je tzv. obsahová sieť. Sú to najnavštevovanejšie české webové stránky [13] a partnerské webové stránky. Reklama sa na týchto weboch zobrazuje podľa obsahu danej stránky. Napríklad ak sa stránka venuje automobily, tak zobrazenou reklamou budú napr. autobazáre, náhradné diely a pod.

Radenie inzerátu ovplyvňuje nastavená maximálna ponúknutá cena za preklik a takzvaný koeficient kvality [14]. Druhý parameter sa počíta pre vyhľadávaciu a obsahovú sieť rôzne. Vo vyhľadávacej sieti je braná do úvahy relevantnosť kľúčového slova v inzeráte ku konkrétnej požiadavke užívateľa, ktorú použije vo vyhľadávaní. V obsahovej sieti sa kladie dôraz na súvislosť reklamy s tematickým zameraním konkrétnej stránky partnerského webu. Ďalší parameter, ktorý ovplyvňuje radenie inzerátu je tzv. miera preklikov. Tá vyjadruje pomer medzi počtom preklikov a počtom zobrazení inzerátu. Teda udáva s akou pravdepodobnosťou si na danú kampaň, zostavu, inzerát alebo kľúčové slovo užívateľa kliknú.

PPC - Google AdWords

Princíp fungovania tejto reklamy je veľmi podobný tomu, ktorý sa nachádza v SKlik. Reklama sa opäť zobrazuje na stránkach s výsledkami vyhľadávania, ale taktiež aj na webových stránkach, ktoré sú vlastnené spoločnosťou Google alebo na partnerských stránkach, ktoré spolu vytvárajú tzv. reklamnú sieť. Umiestňovanie reklamy je automatické podľa kontextu webu a kľúčových slov. AdWords umožňujú pri vytváraní reklamy definovať stránky reklamnej siete, kde má byť reklama umiestnená.

Radenie reklám je takisto ako pri SKlik nástroji ovplyvnený ponukou, ktorá vyjadruje koľko je inzerent ochotný investovať do reklamy a skóre kvality. Skóre ovplyvňuje kvalita reklám inzerenta, kľúčových slov a webových stránok. Podľa toho, kde sa reklama zobrazuje a podľa zvoleného typu cielenia sa vzorec na výpočet hodnotenia reklamy môže meniť, ale vždy sú najpodstatnejšie dva parametre a tými sú ponuka a skóre kvality. Ponuka je opäť ako v prípade SKliku maximálna cena, ktorú inzerent ponúka za preklik, teda cenu, ktorú inzerent chce zaplatiť pri kliknutí na reklamu. Skóre kvality ovplyvňuje relevantnosť použitých kľúčových slov, kvalita vstupnej stránky, miera prekliku a niekoľko ďalších faktorov, ktoré sú však obchodným tajomstvom [4].

5.2.2 Pútač

Možno častejšie používaným slovom je banner, ktorý sa umiestňuje na stránku. Väčšinou ide o reklamu obdĺžnikového alebo štvorcového tvaru. Dnes bannery okrem textu obsahujú animácie, videá bez zvuku alebo so zvukom a v poslednej dobe vznikajú interaktívne bannery, kde užívateľ pohybom myši alebo klávesnicou splní istú úlohu a svoju interakciu

potvrdí kliknutím alebo potvrdením klávesou enter a tým sa dostane na webové stránky, pre ktoré bol banner vytvorený.

5.2.3 Priamy marketing

Ďalšou formou internetového marketingu je priamy marketing. Ide o priame oslovenie zákazníkov, najznámejšou formou tohto typu marketingu je tzv. „newsletter“ teda email odoslaný na účty registrovaných užívateľov, ktorí súhlasili so zasielaním noviniek na svoj emailový účet.

5.2.4 Sociálne siete

Dnes väčšina firiem pochopila, ako je dôležité budovať povedomie o značke a byť v bezprostrednom kontakte so zákazníkmi, či už existujúcimi alebo potenciálnymi. Sociálne siete poskytujú priestor pre vytvorenie sympatií k značke, budovanie dôvery prípadne pre získavanie spätnej väzby. Je možné otestovať obľúbenosť produktu, napríklad jednoduchým zobrazením produktu a požiadaním zákazníkov o hodnotenie prípadne len o obyčajné „lajknutie“.

Je možné predstavovať produkt len určitému segmentu ľudí, podľa veku, regiónu, záujmov a teda cieľiť reklamu len na skupinu ľudí, pre ktorých bude produkt najviac vhodný[19].

5.2.5 Advergaming

Ďalšou formou marketingu je advergaming. Vznikol zložením dvoch anglických slov advertisement and game a voľne preložené to znamená „propagácia pomocou hry“. Ide o nástroj, kedy je naprogramovaná hra, ktorá zábavnou formou distribuuje reklamný odkaz.

Opisom jednotlivých nástrojov online marketingu bol dokončený opis teoretických poznatkov, ktoré s webovou analytikou, atribúciou a marketingom súvisia. Tieto poznatky budú aplikované pri vytváraní nástroja, ktorý je cieľom projektu.

Nasledujúca kapitola popisuje požiadavky na nástroj, ktorý bude umožňovať z nameraných dát interakcie užívateľa s webovou stránkou rozdeliť konverziu podľa atribučných modelov medzi kanály, ktoré sa na konverzii podieľali.

Kapitola 6

Požiadavky a návrh aplikácie

Predchádzajúca kapitola bola poslednou, ktorá popisovala teoretické znalosti, ktoré budú vo zvyšných častiach textu pretransformované na nástroj, ktorý je cieľom projektu.

Táto kapitola sa venuje návrhu nástroja. V úvodnej časti sú popísané funkčné aj nefunkčné požiadavky na aplikáciu, z ktorých budú potom identifikované prípady použitia. V nadväzujúcich častiach sú tieto požiadavky pretvorené do návrhu nástroja. Ten je rozčlenený na komponenty aplikácie, ktorým je priradená funkčnosť a vzájomná závislosť.

6.1 Požiadavky

Pri popise nástroja boli na nástroj kladené dva typy požiadaviek. Prvými sú požiadavky funkčné, ktoré popisujú správanie nástroja a druhými sú nefunkčné požiadavky, ktoré zahŕňujú všetky ostatné požiadavky či už na fyzické zariadenia, použité nástroje, implementačný jazyk atď.

6.1.1 Funkčné požiadavky

Popis funkcionality tohto systému je možné rozdeliť do dvoch oblastí, na meráciu a zobrazovaciu. Meracie požiadavky sa zameriavajú na časť systému, ktorá sa bude zaoberať zberom a ukladaním dát o návštevníkoch. Požiadavky v tejto oblasti sú nasledovné:

- identifikovať návštevníka,
- meranie zobrazenia stránok,
- zaznamenávanie obsahu,
- meranie návštev,
- zaznamenávanie zdrojov požiadaviek,
- definovanie cieľov,
- meranie konverzie,
- zaznamenávanie kampaní.

Druhou skupinou funkčných požiadaviek sú požiadavky pre zobrazenie analyzovaných dát. Požiadavky v tejto oblasti sú nasledovné:

- analyzovanie dát,
 - metrika konverzie,
 - aplikovanie atribučného modelu,
 - analyzovanie dát v istom časovom intervale,
 - rozdelenie metriky podľa dimenzie,
 - filtrovanie dát pri analýze,
- vytvorenie vizualizácie analyzovaných dát,
 - zobrazenie metriky po aplikovaní atribučného modelu,
 - porovnanie dvoch atribučných modelov po aplikácii na metriku.

6.1.2 Nefunkčné požiadavky

Druhá skupina požiadaviek sa zameriava na obmedzenia, ktoré sú kladené na použitú platformu, implementačný jazyk. Tieto požiadavky je možné zhrnúť do nasledujúceho zoznamu:

- základom meracieho komponentu bude nástroj Open web analytics,
- k zberu dát bude dochádzať na webovom serveri,
- dáta sa budú ukladať do systému ElasticSearch,
- pre analýzu a zobrazenie dát sa využije nástroj platformy Pentaho.

6.1.3 Prípady použitia

Na základe špecifikovaných požiadaviek je možné vytvoriť prípady použitia vyvíjaného nástroja. Zo špecifikácií vyplývajú traja aktéri a niekoľko základných akcií, ktoré musí systém spĺňať.

Prvým aktérom v nástroji je *návštevník*, ktorý nepriamo odovzdáva informácie nástroju a je to jediná jeho akcia, ktorú vykonáva v systéme.

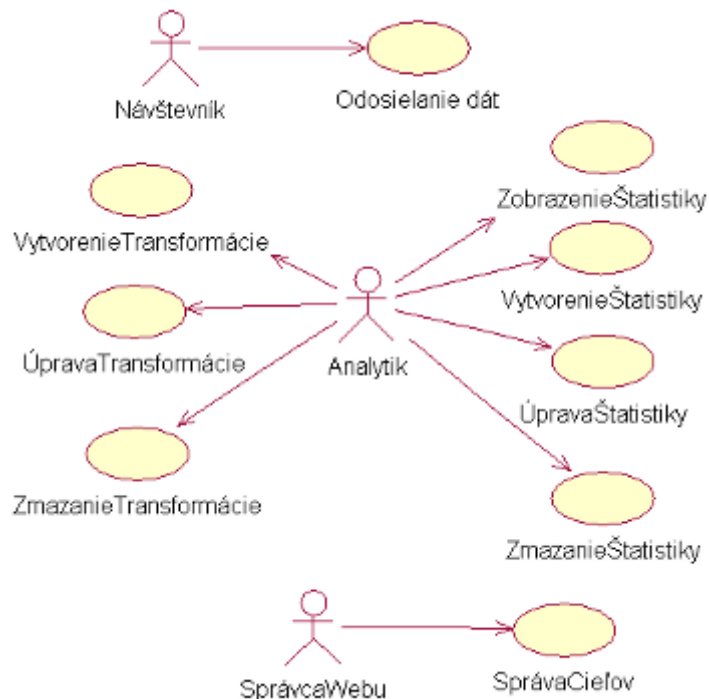
Druhým aktérom je *analytik*. Ten môže vytvárať, upravovať a mazať transformácie dát. Tieto transformácie sa potom využívajú pri zobrazovaní dát. Pomocou transformácie je možné získať dáta a analyzovať ich. Na základe transformácie je možné vytvoriť rôzne štatistiky, o ktoré sa stará práve analytik.

Posledným aktérom, ktorý v systéme existuje, je *správca webu*, ktorý umožňuje definovať ciele webovej stránky.

Po popísaní prípadov použitia je potrebné rozdeliť nástroj na jednotlivé komponenty, ktoré sa podieľajú na zbere, analýze a vizualizácii dát. V nasledujúcej časti je predstavený diagram komponentov aj s jeho popisom.

6.2 Komponenty aplikácie

Diagram komponentov posluží pre grafické zobrazenie komponentov a na modelovanie závislostí medzi jednotlivými komponentami. Na obrázku 6.2 je diagram, ktorý popisuje vytváraný nástroj. Ako prvý komponent v nástroji sa vyskytuje webový prehliadač. Pomocou neho pristupuje užívateľ k webovému serveru, ktorý je tvorený dvomi komponentmi



Obrázek 6.1: Prípady použitia vyvíjaného nástroja

— webovou aplikáciou a trackerom. Webová aplikácia môže byť obecná webová aplikácia, ktorá je dostupná na webovom serveri. V prípade, že si užívateľ vyžiada obsah webovej aplikácie a tento obsah má byť meraný a zaznamenaný ako aktivita, tak túto informáciu predá meraciemu kódu. Webová aplikácia informuje tracker, o ktorú webovú aplikáciu ide a môže mu predáť ďalšie dodatočné informácie akými sú napríklad detaily o obchodných transakciách v internetovom obchode, informácie o akcii, ktorú užívateľ vykonal a iné.

V podstate nie je obmedzené, či tracker patrí vždy len jednej webovej aplikácii. Je možné mať niekoľko webových aplikácií a pre ne používať jeden merací komponent.

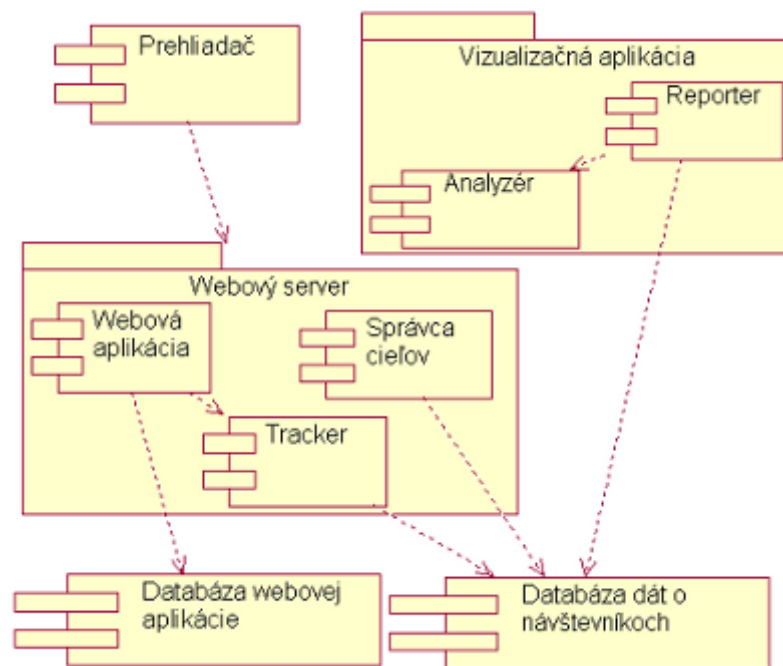
Úlohou trackera je spracovávať v systéme požiadavky, ktoré prichádzajú z webovej aplikácie, na zaznamenanie aktivity návštevníka. Po spracovaní týchto požiadaviek odošle tracker dáta do nástroja pre perzistentné uloženie dát v diagrame, označeného ako databáza dát o návštevníkoch.

Druhá databáza v diagrame, označená ako databáza webovej aplikácie, slúži na zdôraznenie faktu, že dáta pre webovú analytiku nie sú závislé na dátach webovej aplikácie.

Posledný komponent, ktorý sa využíva pri meraní dát, je správca cieľov. Tento komponent umožňuje pre každú stránku definovať obmedzené množstvo cieľov, ktoré pri splnení znamenajú pre stránku konverziu.

Tento popis patrí meracej časti aplikácie. Druhá časť aplikácie slúži na analýzu dát a zobrazovanie výsledkov. Do tejto časti patrí vizualizačná aplikácia, ktorú je možné rozdeliť do dvoch komponentov. Prvý komponent označený ako analyzátor, slúži pre komunikáciu s databázou a spracovanie dát pred ich zobrazením.

Druhý komponent potom dáta od analyzátoru prijme a zobrazuje v štatistikách. Reporter slúži na vytváranie štatistík, ich zobrazovanie a úpravu. Pri zobrazovaní štatistík je možné meniť parametre podľa používanej štatistiky.



Obrázek 6.2: Diagram komponent

Problém určovania atribúcie kampaní je možné rozdeliť do troch čiastkových problémov. Prvým je získanie a zber dát. Nástroj získa informácie o návštevníkovi a spracuje ich. Dáta návštevník o sebe "posiela" tým, že v prehliadači otvorí stránku, vygenerovanú webovou aplikáciou. Táto aplikácia môže pomocou jednej z metód na meranie získať informácie o návštevníkovi, na základe požiadavky, ktorá bola vytvorená. Tieto metódy spracovania sú popísané v sekcii 6.2.1.

Druhý problém, s ktorým sa nástroj musí vysporiadať, je uloženie získaných a predspracovaných dát. V sekcii 7.1.2 je popísaný dátový model a nástroj pre ich uloženie.

Posledný problémom nástroja je analýza dát a vizualizácia analyzovaných dát. Tento problém popisuje sekcia 6.2.3.

6.2.1 Meranie dát

Všeobecne poskytuje užívateľ webovej aplikácii dáta jeho aktivity tak, že pomocou prehliadača vytvorí požiadavku na server. Keďže má aplikácia zbierať informácie o jednotlivých návštevníkoch, tak je vhodné návštevníkov identifikovať. V úvode práce sú spomenuté dve metódy identifikácie pomocou IP adresy a pomocou jednoznačného identifikátora uloženého v trvalom úložisku. Druhý spôsob je presnejší a preto je použitý aj v projekte. Táto identifikácia sa vytvára podľa zvolenej metódy. Tie sú popísané v nasledujúcej sekcii. Na záver fázy merania dát dôjde k spracovaniu tých nazbieraných dát a k ich uloženiu, o ktorom hovorí sekcia 6.2.2

Nástroj, ktorého základ je použitý pre meranie dát v rámci projektu, umožňuje meranie dát o návštevníkoch dvomi spôsobmi:

- meranie na klientovi

- meranie na serveri

6.2.2 Uloženie dát

Získané a spracované dáta sú ukladané do perzistentného úložiska. Tým je Elasticsearch, ktorý poskytuje ukladanie a dostupnosť dát v reálnom čase, tiež vysokú dostupnosť a distribúciu dát [1].

6.2.3 Analýza a zobrazenie dát

Poslednou časťou, ktorú projekt rieši, je zobrazenie a analýza dát. Komponent, ktorý sa o túto časť stará bude komunikovať len s úložiskom dát. Neprichádza už do styku s meracím komponentom. Tým dochádza k nízkej závislosti medzi meracou a zobrazovacou časťou. Komponent pre analýzu a zobrazenie dát získava pri procese ETL¹. Proces zahŕňa získanie dát z úložiska, v prípade rozšírenia úložísk by mohlo byť viac a rôzneho typu, okrem dát napríklad z Elasticsearch, to môžu byť zdroje dát akými sú MySQL databázy, súbory a iné. Ďalším krokom v procese je vyčistenie dát, napríklad doplnením chýbajúcich hodnôt, odstránením nezmyselných dát, preformátovaním dát do spoločného formátu. Posledným krokom je nahrať dáta pre analýzu a zobrazenie.

¹ ETL - skratka z anglických slov Extract, Transform, Load

Kapitola 7

Implementácia

V predchádzajúcej kapitole sú popísané požiadavky, ktoré má nástroj spĺňať. Okrem nich sú tam predstavené jednotlivé komponenty a ich vzájomná prepojenosť.

Táto kapitola sa venuje popisu implementácie. Tú je možné rozdeliť do fáz, tak ako postupne prebiehali. V každej fáze sú spomenuté ciele, problémy a nakoniec úlohy, ktoré sa v danej fáze riešili.

V prvej fáze sa rozhodovalo medzi potenciálnymi riešeniami zberu dát. Nakoniec sa vybral jeden spôsob, ktorý bol implementovaný. V tejto fáze sa okrem merania dát ukladali namerané dáta, a preto je v jednej podkapitole predstavený model dát.

V druhej fáze sa prepojila databáza s nástrojom na spracovanie dát a analýzu, ktorý je súčasťou balíka Pentaho. Kapitola popisuje spracovanie a získavanie dát.

Posledná fáza je venovaná vytvoreniu vizualizácie dát. Dáta je možné zobrazovať pomocou webového rozhrania alebo desktopovej aplikácie. Nakoniec bolo vybrané desktopové rozhranie, ktoré je v krátkosti predstavené v závere kapitoly.

7.1 Implementácia zberu dát

Jednou z požiadaviek na systém bolo využiť merací kód z nástroja Open web analytics. Tento nástroj je implementovaný v jazyku PHP a umožňuje dáta ukladať do MySQL databázy. Avšak projekt má dáta ukladať do úložiska Elasticsearch.

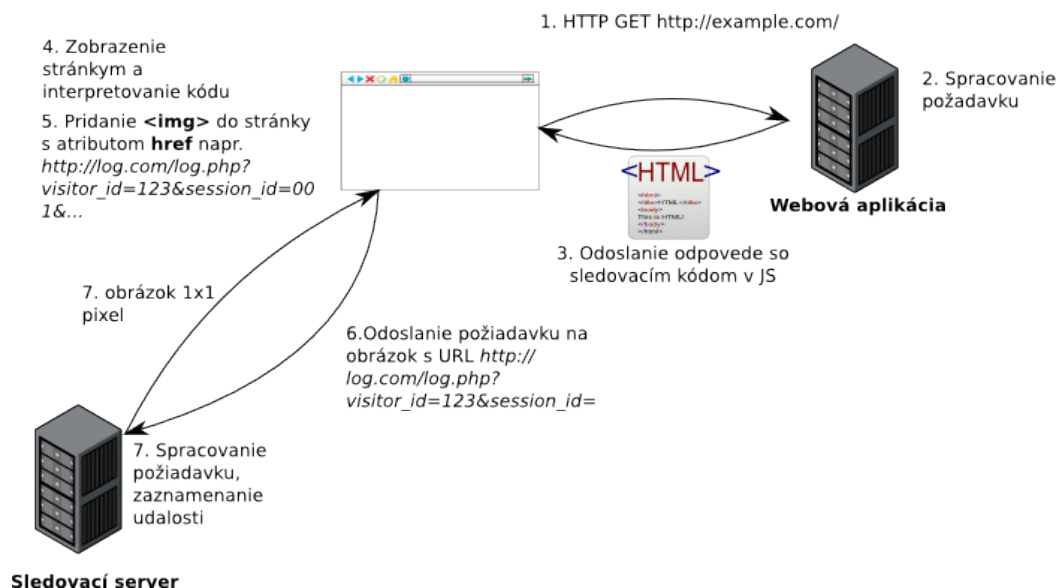
Preto bolo jedným z problémov ukladania dát prepojenie týchto dvoch nástrojov. Ešte skôr ako bude popísané riešenie tohto problému, budú predstavené iné problémy, ktorými sú spôsoby ako merať dáta pomocou nástroja Open web analytics a akým spôsobom namerané dáta spracovať.

Open web analytics umožňuje dáta merať dvomi spôsobmi:

- meranie na klientovi,
- meranie na serveri.

Meranie na klientovi sa deje pomocou kódu napísaného v jazyku Javascript. Obrázok 7.1 znázorňuje komunikáciu a priebeh merania dát.

V prvom kroku zadá užívateľ v prehliadači adresu, ktorej obsah chce zobraziť. Požiadavka je odoslaná na server, na ktorom je webová aplikácia. V druhom bode spracuje webový server požiadavku a vygeneruje obsah ako dokument HTML, ktorý obsahuje vložený sledovací kód. V treťom kroku odpovedá webový server klientovi (webovému prehliadaču). V ďalšom kroku je zobrazený obsah stránky a sledovací kód interpretovaný. V tomto kroku



Obrázek 7.1: Meranie pomocou klientského Javascriptu

dochádza k získaniu dát o návštevníkovi. Vytvorí sa jeho identifikácia, ktorá je uložená do prehliadača pomocou metódy cookies. Okrem identifikácie návštevníka sa vytvára napríklad identifikácia sedenia, ktorá zodpovedá návšteve, tak ako to bolo uvedené pri definovaní termínov. Po získaní dát je ďalším krokom prenos dát na sledovací server. Obrázok 7.1 obsahuje dva servery, avšak webový server môže byť zároveň aj sledovacím serverom.

Dáta sú prenášané na server tak, že sledovací kód dynamicky vloží do zobrazeného dokumentu značku pre obrázok a ako zdroj obrázku nastaví URL pre sledovací server. Táto URL ako dotazovací reťazec obsahuje dáta, ktoré sú zozbierané na klientovi. Napríklad je to už spomínaná identifikácia návštevníka, sedenia, URL sledovanej stránky, URL stránky, z ktorej sa prišlo na sledovanú stránku, aká kampaň bola použitá pri upútaní návštevníka a iné.

Tým, že sa do stránky pridá značka pre obrázok a nastaví sa zdroj obrázku, prehliadač automaticky spustí nahrávanie obrázku z adresy určenej URL. Tento krok je označený v obrázku 7.1 číslom šesť.

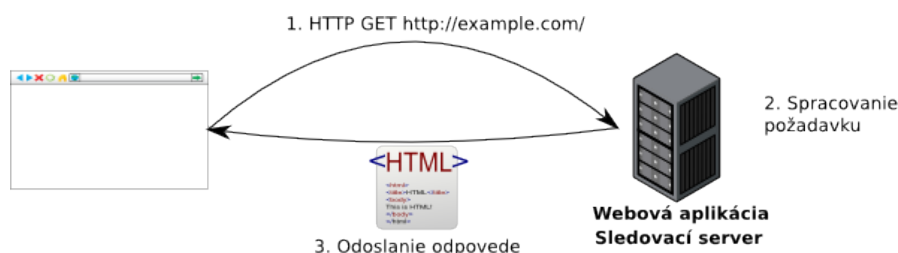
V siedmom kroku rozloží sledovací server dotazovací reťazec a získané dáta dodatočne spracuje. Následne sa dáta uložia a klientovi je odoslaná odpoveď. Odpoveďou je obrázok s rozmermi 1x1 pixel.

Týmto je meranie dát pomocou klienta ukončené. Pri tomto spôsobe merania je možné merať aj dáta, ktoré sú spojené s vstupno-výstupnými zariadeniami, napríklad pohyby myši, kliknutia na jednotlivé časti dokumentu a iné. Tieto dáta pomáhajú lepšie pochopiť, ako pracuje užívateľ s danou stránkou a podľa toho sa môže optimalizovať jej rozvrhnutie alebo spôsob používania.

Avšak použitie Javascriptu má nevýhody v tom, že kód musí byť distribuovaný na klientov, čo zvyšuje prevádzku servera, aj keď môže byť využité nahrávanie z vyrovnávacej pamäte. Ďalším problémom je prehliadač. Nikde nie je zabezpečené, že má každý webový prehliadač povolený Javascript. A teda by dochádzalo k tomu, že užívateľ nebude zaznamenaný ako návštevník. Tretím problémom spojeným s týmto meraním je kompatibilita prehliadačov a kódu v Javascripte. Kód v Javascripte je nutné optimalizovať na čo najväčší

počet prehliadačov, aby projekt meral výsledky čo najpresnejšie. Posledným problémom, ktorý je spojený s týmto spôsobom merania, sú mobilné zariadenia. Prehliadač a Javascript sú opäť slabými článkami, ktoré by mohli spôsobiť problémy pri meraní tým, že nie je povolený Javascript alebo, že prehliadače v zariadeniach podporujú klientský Javascript, ktorý je nekompatibilný s meracím kódom, ktorý by bolo nutné zistiť a optimalizovať.

Meranie na serveri je zobrazené na obrázku 7.2. Prvotná odlišnosť spočíva v tom, že webový server musí okrem webovej aplikácie obsahovať aj kód na meranie a spracovanie dát. Predtým bol kód na meranie prenášaný webovému prehliadaču a kód na spracovanie dát mohol byť uložený aj na inom serveri ako bol webový server s aplikáciou. Teraz je podobný merací kód, ale v jazyku PHP, uložený na serveri s webovou aplikáciou. Ide o kód nástroja Open web analytics odľahčený o jeho grafické užívateľské rozhranie. Prvý krok



Obrázek 7.2: Meranie pomocou servera

merania sa zhoduje s prvým krokom v predchádzajúcej metóde. Druhý krok je už odlišný. V ňom dochádza k spracovaniu požiadavky na strane servera, teda vygeneruje sa obsah stránky. Pri generovaní dôjde k spusteniu sledovacieho komponentu, ktorý získa na základe požiadavky na obsah potrebné dáta, ktoré následne uloží do zvoleného úložiska.

V tomto spôsobe je identifikácia návštevníka generovaná na strane servera, ale ako úložisko sa opäť využije webový prehliadač s jeho technológiou - cookies.

V poslednom kroku sa do webového prehliadača prenáša len obsah požadovanej stránky. Webový prehliadač nenačítava žiadny kód a logicky nedochádza ani k jeho interpretácii.

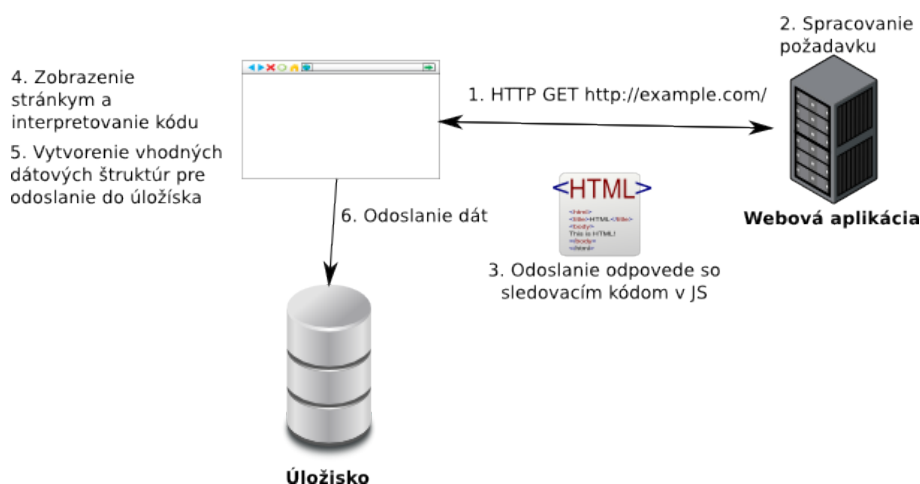
Jednou z výhod tohto prístupu je, že meranie sa deje na serveri a nie je nutné riešiť kompatibilitu kódu s jednotlivými prehliadačmi. Ďalšou výhodou je, že merací kód zostáva skrytý pred návštevníkom.

Tento spôsob prináša nevýhodu v tom, že nie je možné merať interakciu užívateľa s prehliadačom, tak ako je tomu pri prvom spôsobe merania. Avšak projekt nevyžaduje merať túto interakciu medzi návštevníkom a prehliadačom a teda môže byť zanedbaná.

Problém merania a ukladania dát môže byť vyriešený tromi rôznymi spôsobmi. Tie sú popísané nasledovne:

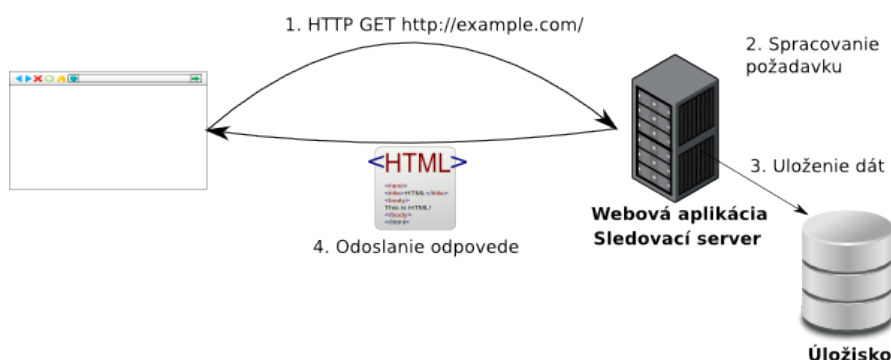
- **Spracovanie na klientovi** - tento prístup predpokladá implementovanie celej logiky spracovania dát pomocou kódu v Javascripte. Základné zbieranie sa deje rovnako, ako to je popísané v úvode tejto kapitoly. Získa sa identifikácia pre návštevníka, návštevu a podobne. Ďalší krok, ktorý je odlišný od toho pôvodného, je, že dáta musia byť štruktúrované, tak aby mohli byť odoslané priamo do úložiska. Nakoniec dochádza k odoslaniu dát do úložiska. Výhodou tohto prístupu je, že pri meraní nedochádza ku komunikácii s webovým serverom, ktorý by mal potom ešte spracovávať dáta a uložiť ich do úložiska. Po výskúšaní tohto prístupu sa zistilo, že merací kód sa výrazne zväčšuje a dosahoval približne rovnakú veľkosť ako bola samotná stránka

v HTML kóde. Podľa typu dát, ktoré sa ukladali do úložiska, dochádzalo aj k vzrastajúcemu množstvu požiadaviek na server, čo zaťažovalo klienta a pri testovaní nebolo toto meranie použiteľné na mobilných zariadeniach.



Obrázek 7.3: Spracovanie na klientovi bez pozornosti servera

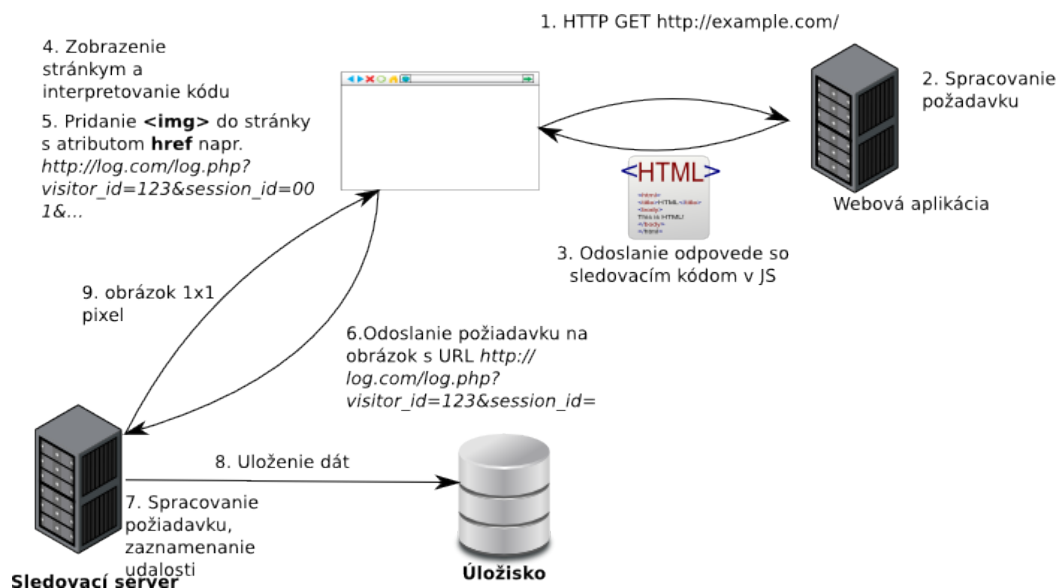
- **Spracovanie na serveri** - druhý prístup merania je popísaný a znázornený na obrázku 7.2. Na obrázku nie je znázornená komunikácia s úložiskom, takže je, pre úplnosť, doplnená na obrázku 7.4. Tento prístup má nevýhody, ktoré sú popísané vyššie. Napriek tomu nie je meranie závislé na niekoľkých platformách (rôzne prehliadače, rôzne operačné systémy), ale vždy len na tej, na ktorej server beží.



Obrázek 7.4: Spracovanie len na serveri

- **Meranie na klientovi a spracovanie na serveri** - posledný prístup popisuje obrázok 7.1. Opäť, kvôli úplnosti, je na obrázku 7.5 doplnená komunikácia s úložiskom.

Pre tento projekt bol použitý druhý spôsob merania dát. Ten poskytuje najmenšiu závislosť na platforme, dostatočné množstvo dát, ktoré je nutné pre dosiahnutie cieľa, o ktorý sa projekt usiluje.



Obrázek 7.5: Meranie na klientovi spracovanie na serveri

7.1.1 Implementácia komunikácie medzi Elasticsearch a OWA

Po výbere metódy merania dát bolo nutné vyriešiť problém s prepojením Open web analytics a Elasticsearch. Open web analytics využíva pre komunikáciu s databázou ovládač, ktorý je objektom triedy *owa_db*. Tento ovládač plní v nástroji dve úlohy:

- Konštruktér SQL dotazov - pomocou volania metód dochádza k zbieraniu parametrov SQL dotazu. Tieto metódy predstavujú jednotlivé klauzuly SQL dotazu napr. `select`, `where`, `from`, `order`, `group by` atď.
- Komunikátor s databázou - po nastavení parametrov SQL dotazu dôjde k jeho vygenerovaniu a vykonaniu na strane databázového servera.

To znamená, že úlohou pri implementovaní komunikácie bolo vhodným spôsobom vytvoriť druhú úlohu ovládača.

Na základe parametrov SQL dotazu bolo nutné vytvárať dotazy pre Elasticsearch.

ElasticSearch komunikuje s klientmi pomocou protokolu HTTP a dáta, ktoré sa posielajú serveru sú v otvorenej forme a vo formáte JSON. Pre zjednodušenie práce pri vytváraní dotazov bol využitý klient s názvom *Elastica*. Tento klient umožňuje vytvárať dotazy pomocou ORM¹ rozhrania.

Projekt implementuje mapovanie len istej podmnožiny všetkých parametrov, ktoré je možné v databázovom ovládači nastavovať. Ide len o tie parametre, ktoré sa využívajú pri zbere a ukladaní dát. Ide o nasledujúce parametre pre klauzule:

- `select` - parametre pre projekciu,
- `where` - parametre pre selekciu, podpora rovnosti v klauzule `where`,
- `from` - výber tabuľky,

¹V angličtine Object relational mapping v tomto zmysle ide o mapovanie fragmentov dotazu na objekty

- insert - vkladanie dát,
- update - zmena dát.

V rámci projektu boli implementované nasledovné agregáčn  funkcie:

- suma,
- po et,
- maximum.

Tieto agregáčn  funkcie s  mapované na tzv. facets v Elasticsearch. Facets plnia v datab ze presne rovnak  funkciu ako agregáčn  funkcie v MySQL.

ElasticSearch implicitne neumo ňuje okam ite pracovať s pr ve ulo en mi d tami. V podstate vznik  pr bl m, kedy aj po  spešnom odoslan  a spracovan  dotazu nie s  zmeny okam ite viditeľn  v nasleduj cich dotazoch. Pri tomto projekte je nutn  vidieť zmeny okam ite, preto e niektor  inform cie s  po ítan  z pr ve ulo en ch d t. Z toho d vodu je nutn  ka d  dotaz, ktor  vykon va zmeny, potvrdzovať, aby boli okam ite viditeľn .

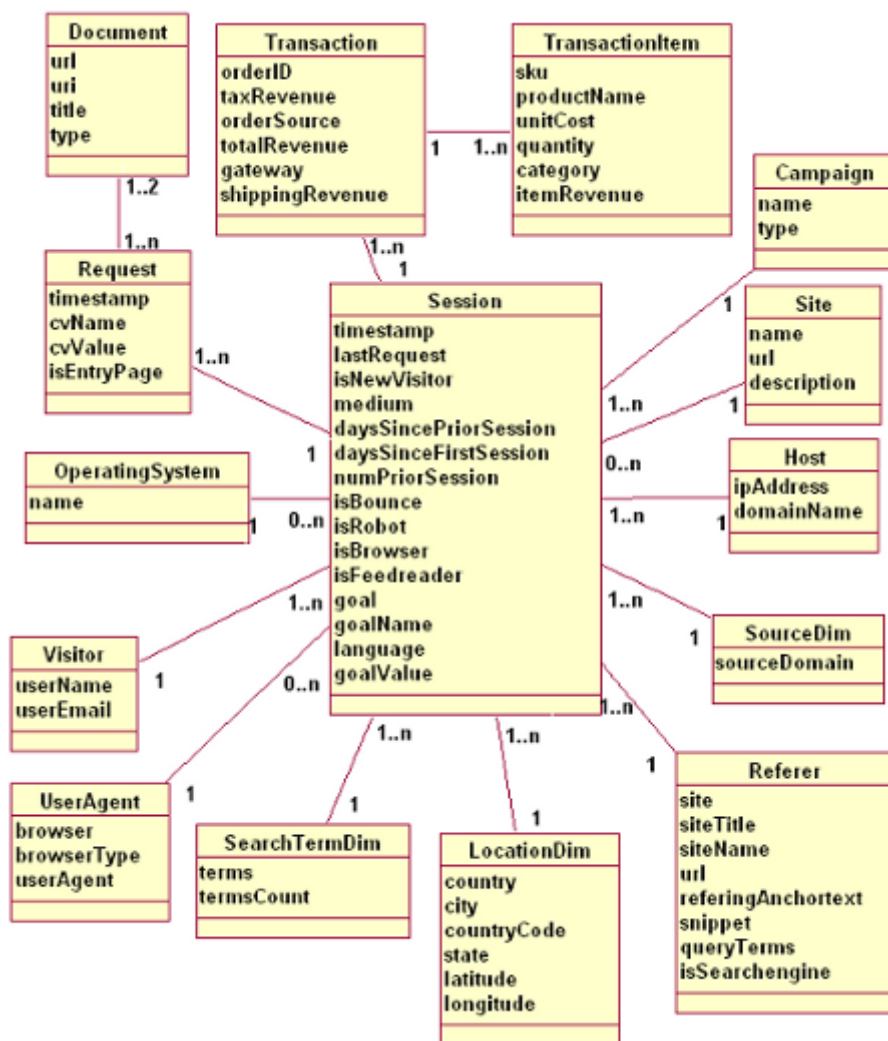
Po opise komunik cie medzi Elasticsearch a Open web analytics je v nasleduj cej  asti predstaven  d tov  model.

7.1.2 Datab zov  model

Na z klade po iadaviek z predch dzajúcej kapitoly bol vytvoren  model, ktor  reprezentuje ukladan  d ta.

Na obr zku 7.6 je mo n  vidieť jednotliv  triedy a ich vz ahy, ktor  sa pri meran  d t ziskavaj . V nasleduj cej  asti je vysvetlen  v znam jednotliv ch tried a ich logick  prepojenie:

- *Site* - reprezentuje webov  str nku, ku ktorej sa vz ahuj  v etky ostatn  nameran  d ta. Atrib ty tejto triedy s :
 - *site* - meno webovej str nky,
 - *url* - verejn  URL, na ktorej je webov  str nka dostupn ,
 - *description* - kr tky popis webovej str nky.
- *Session* - reprezentuje sedenie alebo lep ie povedan  n v tevku webovej str nky. Inform cie, ktor  potrebuje n stroj o n v tev ch vedieť, s  nasledovn :
 - *timestamp* -  as kedy n v steva za ala,
 - *lastRequest* -  as kedy v r mci danej n v stevy vytvoril u  v teľ po iadavku na server.
 - *isNewVisitor* - zna ka,  i ide o n v tevku, keď pri iel u  v teľ prv krát na webové str nky,
 - *medium* - ozna enie skupiny zdrojov, z ktor ch sa u  v teľ dostal na str nky.
 - *daysSincePriorSession* - po et dn  od predch dzajúcej n v stevy,
 - *daysSinceFirstSession* - po et dn  od prvej n v stevy,
 - *numPriorSessions* - celkov  po et predch dzaj cich n v stev,



Obrázek 7.6: Konceptuálny model

- *isBounce* - značka pre označenie neúmyselnej návštevy, v prípade, že užívateľ ukončil svoje prehliadanie už na prvej stránke,
 - *isRobot* - značka pre označenie, či ide o robota (vyhľadávač, malware), ktorý navštívil webovú stránku,
 - *isBrowser* - značka pre označenie, či ide o prehliadač, ktorý komunikuje s webovou stránkou,
 - *language* - informácie o lokalizácii prehliadača,
 - *isFeedReader* - značka pre označenie, či ide o aplikáciu na čítanie RSS kanálov,
 - *goal*, *goalName*, *goalValue* - slúžia na reprezentáciu dát spojených s dosahovaním konverzie.
- *Visitor* - reprezentuje návštevníka webovej stránky. Atribúty sú nasledovné:
 - *userName* - užívateľské meno,

- *userEmail* - adresa elektronickej pošty.
- *Request* - reprezentuje požiadavku návštevníka na zobrazenie obsahu jednej stránky. Každá návšteva obsahuje aspoň jednu požiadavku na zobrazenie stránky. Každá požiadavka má nasledovné atribúty:
 - *timestamp* - čas kedy k požiadavke došlo,
 - *isEntryPage* - označenie, že požiadavka bola prvá v rámci návštevy,
 - *customValueName*, *customValueValue* - nastavenie doplnkových informácií pre danú požiadavku.
- *Document* - reprezentuje stránku, ktorá bola zobrazená návštevníkovi na základe jeho žiadosti. Jedna požiadavka môže obsahovať dva dokumenty. Jeden je aktuálne zobrazený a druhý označuje predchádzajúci dokument. Atribúty viazané ku každému dokumentu sú:
 - *URL* - verejná adresa dokumentu,
 - *URI* - umiestnenie dokumentu na serveri,
 - *title* - titulok stránky,
 - *type* - typ stránky.
- *OperatingSystem* - reprezentuje operačný systém, z ktorého sa návštevník pripojil k serveru a má jediný atribút:
 - *name* - meno operačného systému.
- *UserAgent* - reprezentuje prehliadač, ktorý bol pri prístupe k serveru použitý. Počas jednej návštevy sa použije práve jeden prehliadač. A naopak, prehliadač sa môže použiť počas niekoľkých návštev. Atribúty spojené s prehliadačom sú:
 - *userAgent* - úplná informácia o klientovi obsahuje operačný systém, prehliadač, meno jadra prehliadača a jeho verziu,
 - *browser* - meno prehliadača,
 - *browserType* - typ prehliadača.
- *SearchTermDim* - reprezentuje reťazec, ktorý bol zadaný do vyhľadávacieho a návštevník potom na základe výsledkov vo vyhľadávacom prišiel na sledovanú webovú stránku. Atribúty sú:
 - *terms* - výraz zadaný vo vyhľadávacom,
 - *termsCount* - koľkokrát bol tento výraz použitý vo vyhľadávacom návštevníkmi a tí potom navštívili webové stránky.
- *LocationDim* - reprezentuje lokalitu, z ktorej sa užívateľ pripája. Podobne ako pri prehliadači aj tu je vzťah lokality a návštevy jedna ku N. Jedna lokalita je súčasťou viacerých návštev a každá návšteva je spojená s jednou lokalitou. S touto informáciou sú spájané nasledovné atribúty:
 - *country* - krajina v ktorej sa návštevník nachádza,
 - *city* - mesto, v ktorom sa nachádza,

- *state* - štát, provincia, kraj, v ktorom sa návštevník nachádza,
 - *longitude* - zemepisná dĺžka,
 - *latitude* - zemepisná šírka.
- *Referer* - reprezentuje zdroj, z ktorého návštevník prišiel na sledované webové stránky. Jedna návšteva môže mať jednu odkazujúcu stránku. Atribúty, ktoré nástroj ukladá v spojení s odkazujúcou stránkou sú:
 - *site* - meno stránky,
 - *siteName* - titulok stránky, na ktorej bol odkaz na sledovanú stránku,
 - *url* - URL odkazujúcej stránky,
 - *queryTerms* - použité termíny pri hľadaní,
 - *referringAnchortext* - text elementu odkazu na odkazujúcej stránke,
 - *snippet* - časť html dokumentu odkazujúcej stránky,
 - *isSearchengine* - označenie, či je odkazujúca stránka vyhľadávačom alebo nie.
 - *SourceDim* - reprezentuje zdroj, z ktorého návštevníci prichádzajú na stránky. Zdrojom môže byť reklamná pošta či konkrétny časopis. V porovnaní s médiom je zdrojom konkrétny kanál. Napríklad reklamná pošta v emailovej schránke je zdrojom a médiom je v takomto prípade email. Nástroj bude zaznamenávať zdroje podľa mena:
 - *name* - názov zdroja.
 - *Host* - reprezentácia zariadenia, z ktorého prišla požiadavka. Atribúty uchovávané pre identifikáciu zariadenia:
 - *ipAddress* - IP adresa,
 - *hostName* - doménové meno zariadenia.
 - *Transaction* - reprezentácia obchodnej transakcie. V rámci jednej návštevy môže zákazník vytvoriť niekoľko transakcií.
 - *orderID* - identifikácia objednávky v elektronickom obchode,
 - *taxRevenue* - celkové daň pri objednávke,
 - *totalRevenue* - celková cena pri objednávke,
 - *shippingRevenue* - celková cena dopravy,
 - *gateway* - použitá platobná brána,
 - *orderSource* - zdroj, ktorý doviedol návštevníka na stránky a počas svojej návštevy vykonal transakciu.
 - *TransactionItem* - reprezentuje jednu položku v transakcii. Každá transakcia môže obsahovať niekoľko položiek, ale jedna položka nemôže byť vo viacerých transakciách.
 - *sku* - identifikácia tovaru v elektronickom obchode,
 - *productName* - názov produktu,
 - *unitCost* - cena za jeden kus produktu,
 - *quantity* - počet kusov produktu v transakcii,

- *category* - kategória produktu,
- *itemRevenue* - celková cena za položku v transakcii, získa sa vynásobením jednotkovej ceny počtom kusov.
- *Campaign* - reprezentuje kampaň, ktorá návštevníka prilákala na webové stránky. Viacero návštev môže vzniknúť na podnet jednej kampane, ale návšteva vzniká na základe jednej kampane.
 - *name* - názov kampane,
 - *type* - typ kampane.

Databázový model je založený na type hviezdy. Hlavnou tabuľkou je tzv. tabuľka faktu. V našom prípade je touto tabuľkou tabuľka pre návštevy alebo požiadavky. Jej hlavným rysom je, že obsahuje cudzie kľúče do iných tabuliek. Tieto sa nazývajú tabuľky dimenzií a detailnejšie popisujú dáta, ktoré sa nachádzajú v tabuľke faktu.

V prípade tohto projektu sú tabuľkami dimenzií tabuľka dokumentu, kampane, prehliadač, operačný systém, zdroje prevádzky, vyhľadávacie termíny, pozícia, hostiteľ, atď.

7.2 Implementácia získavania a analyzovania dát

Po fáze zbierania a ukladania dát prišla na rad fáza spracovania dát. Pri nej bola použitá platforma Pentaho. Tá umožňuje prístup k zozbieraným dátam, ich integráciu a vizualizáciu. V projekte použijeme len časť celej platformy.

Pentaho umožňuje vytváranie súhrnov (štatistík) pomocou:

- webového rozhrania,
- desktopovej aplikácie.

V projekte bol použitý druhý spôsob pre širšiu funkcionality, ktorá bola zistená pri počítačnom testovaní. Desktopová aplikácia nesie názov Pentaho report designer. Slúži na vizualizáciu dát.

Druhým nástrojom, ktorý nebol spomenutý, je nástroj pre ETL proces s názvom Pentaho data integration. Oba nástroje sú súčasťou platformy Pentaho.

ETL proces ako termín vznikol z anglických slov extracting, transforming a loading. Ide o proces, kde dochádza k získaniu dát z externých zdrojov (Extracting). V ďalšom kroku dochádza k transformácii dát získaných v prvom kroku, tak aby vyhovovali potrebám pre ďalšiu prácu. Posledným krokom je nahrať (loading) transformovaných dát do cieľa.

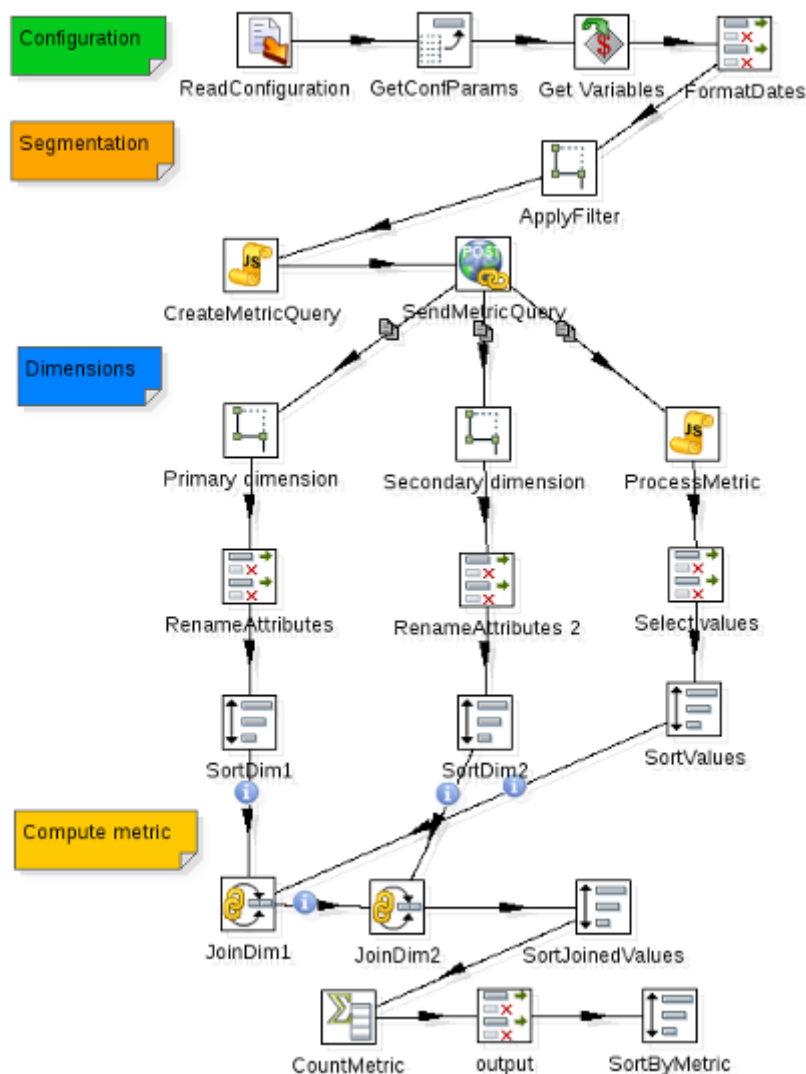
Základným kameňom pri tomto procese je transformácia, ktorá je reprezentovaná acyklickým orientovaným grafom. Jednotlivé uzly grafu sú funkčnými blokmi celého procesu. Tieto bloky sa pri spustení transformácie vykonávajú paralelne. Pokiaľ existuje vstup do daného funkčného bloku, tak je transformovaný na výstup a odoslaný na do ďalšieho funkčného bloku. Toto predstavenie slúži ako základ pre pochopenie transformácie v nástroji pre integráciu dát.

Podľa požiadaviek, ktoré boli stanovené na začiatku projektu, má byť možné vytvárať metriky. V nástroji Pentaho predstavuje počítanie metriky transformácia. Tá je parametrizovaná a je pri nej možné špecifikovať dimenzie, podľa ktorých majú byť rozdelené výsledky a filtre pre výber určitého segmentu dát.

Transformácie pre metriky sa odlišujú tým, akým spôsobom sú počítané. V nasledujúcej podkapitole je predstavená transformácia pre počítanie metrík návštev a počtu zobrazení stránky. Transformácia obsahuje bloky, ktoré môžu byť znovu použité pri vytváraní iných transformácií, pretože plnia funkcionality spoločnú pre všetky transformácie.

7.2.1 Transformácia základných metrík

Na obrázku 7.7 je zobrazená transformácia, ktorá počíta základné metriky. Obecnne je možné transformáciu rozdeliť do štyroch častí.



Obrázek 7.7: Transformácie pre metriku

Prvou časťou je konfiguračná časť, označená na obrázku 7.7 ako *Configuration*. V tejto časti dochádza k načítaniu konfiguračného súboru. Formát dát v tomto súbore podporuje používané konštrukcie, ktoré triedy *java.util.Properties* jazyku Java umožňuje spracovávať. Aktuálne sa zo súboru získavajú URL a meno databázy. Po načítaní súboru a spracovaní nastavení sú v ďalšom kroku získané parametre transformácie. Túto úlohu vykonáva funkčný

blok *GetParameters*. V poslednom bloku konfiguračnej časti potom dochádza k formátovaniu dátumu, ktorý je jeden zo vstupných parametrov transformácie, do formátu použitého pri ukladaní dát.

Každá transformácia obsahuje niekoľko parametrov. Povinnými parametrami pri každej transformácii sú:

- časový interval - obdobie, pre ktoré sa zobrazuje štatistika,
- identifikácia stránky,
- metrika - metrika, ktorá sa zobrazuje.

Okrem povinných parametrov je možné transformácii nastaviť aj nepovinné parametre, ktoré sa líšia od typu transformácie, ale nasledujúce parametre je možné nastaviť pre každú transformáciu:

- primárna dimenzia,
- sekundárna dimenzia,
- parametre filtra.

Druhá časť označená ako *Filter* slúži pre segmentáciu dát. Podľa nepovinných parametrov filtra sa vytvorí dotaz na databázu a získajú sa len dáta, z ktorých sa v ďalšej časti transformácie počítajú metriky.

Parametre, ktoré je možné nastaviť ako parametre filtra, sú:

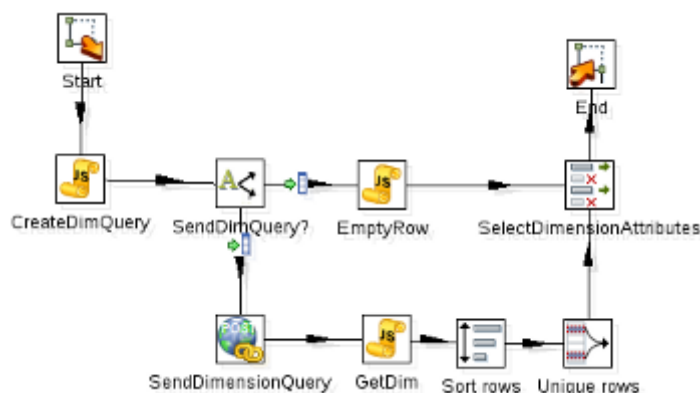
- dimenzia filtra,
- typ filtra,
- hodnota.

V tretej časti sa na základe filtra získajú segmentované dáta. Pokiaľ nie je filter zadaný, tak sa získajú všetky dáta v sledovanom časovom období. Prvým krokom pri získaní dát je vytvorenie dotazu na databázu. Za túto akciu zodpovedá blok *CreateMetricQuery*. Následne je tento dotaz odoslaný pomocou bloku *SendMetricQuery*. Získané dáta sú z databázy skopírované (nie distribuované) do blokov pre získanie dát dimenzie a do časti pre počítanie metriky. Ak boli by dáta distribuované do jednotlivých blokov, tak by došlo k rozdeleniu výstupného toku dát z bloku *SendMetricQuery*. Transformácia chce spracovať výstupný tok zvlášť v každom bloku, preto dochádza k jeho kopírovaniu.

V štvrtej časti sa zisťujú dáta dimenzie. Niektoré dimenzie sú súčasťou tabuľky faktu napr. dátum, počet predchádzajúcich návštev a iné. Niektoré dimenzie sú k tabuľke faktu pridružené cudzím kľúčom a je nutné detaily z takejto tabuľky pripojiť k výsledku. V prípade, že je požadovaná dimenzia v tabuľke detailu, vytvorí sa dotaz, ktorý získa požadované dáta, na základe vstupného toku, ktorý pochádzal z bloku *SendMetricQuery*. V prípade, že je dimenzia súčasťou tabuľky faktu, vkladá sa do výstupu prázdny riadok. Inak by mohlo dôjsť k zastaveniu spracovania dátového toku.

Bloky *Primary dimension* a *Secondary dimension* sú podtransformáciami. O získanie dát dimenzie sa stará práve transformácia, ktorá je zobrazená na obrázku 7.8. Každá podtransformácia musí začínať blokom s názvom *Start* a končiť s blokom s názvom *End*. Potom sa v transformácii vytvára dotaz na databázu. Pokiaľ nie je dimenzia špecifikovaná, pokračuje sa blokom *EmptyRow*, ktorý do toku dát vloží prázdny riadok. V opačnom prípade,

ak bola dimenzia špecifikovaná, je dotaz odoslaný do databázy, zistia sa potrebné dáta, z ktorých sú následne odstránené duplicity. V bloku *SelectDimensionAttributes* dôjde k výberu len tých dát, ktoré boli získané z databázy.



Obrázek 7.8: Podtransformácia dimenzie

V blokoch *RenameAttributes* a *RenameAttributes2* sa premenujú dáta dimenzií.

V blokoch *JoinDim1* a *JoinDim2* sú postupne pripojené dáta dimenzií k dátam metricky pomocou cudzieho kľúča, pod ktorým vystupujú dimenzie v tabuľke faktu.

V bloku *CountMetric* sú dáta zoskupené podľa dát dimenzií a je spočítaný počet riadkov s rovnakými hodnotami pri daných atribútoch. Nakoniec sú v poslednom kroku z toku vybrané dáta dimenzií a metrika. V poslednom bloku potom dochádza k zoradeniu výsledkov podľa hodnôt metricky.

Pri vytváraní transformácie pre metricky museli byť akceptované isté obmedzenia, ktoré prichádzajú s jej použitím. Najlepšie je obmedzenie demonštrovať na príklade. Napríklad nie je možné špecifikovať zoznam dimenzií v jednom parametri a potom v cykle zistiť dáta a pripojiť ich k výsledku. Odporuje to transformácií, ktorá je acyklickým grafom. Čiže pri spustení transformácie dochádza k spusteniu daného uzla v grafe len raz pre daný riadok dát v toku. Preto je nutné špecifikovať maximálne množstvo dimenzií v danej transformácii. Potom je každá dimenzia spracovaná vo vlastnom podstrome grafu.

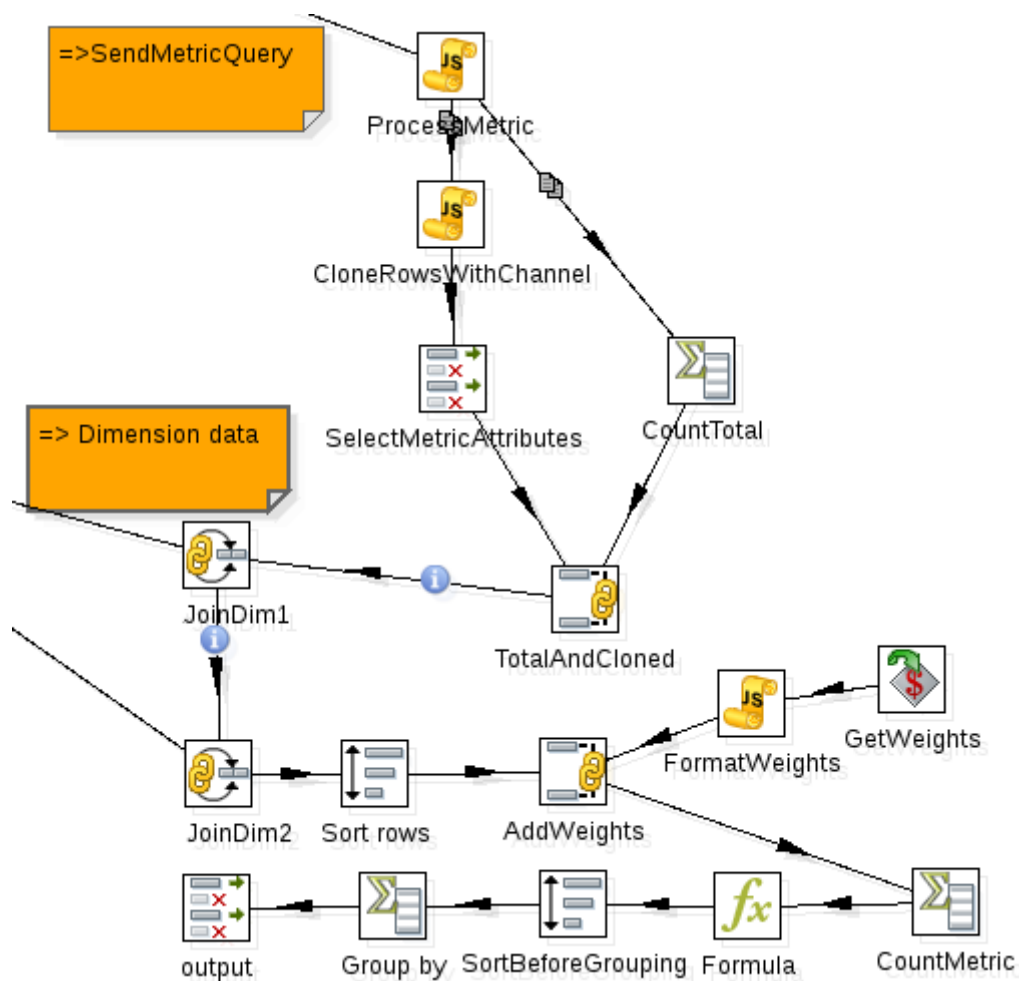
7.2.2 Transformácia s aplikovaním atribučného modelu

Ako základ slúži transformácia popísaná v predchádzajúcej kapitole. Obsahuje konfiguračnú časť, časť dimenzií a filtrovaciu časť. V čom sa líši od tej predchádzajúcej, je časť, ktorá počíta konverziu pri aplikovaní modelu. Na obrázku 7.9 je uvedená len táto časť, dôležitá pre popis výpočtu.

Blok *ProcessMetric* je zhodný s blokom z predchádzajúceho obrázku a od neho dochádza k zmene pri spracovaní dát oproti predchádzajúcej transformácii.

V dátovom toku obsahuje každý riadok atribút, v ktorom je zoznam kanálov, ktoré návštevník použil počas návštevy. Takýto riadok so zoznamom o dĺžke N je transformovaný na N riadkov s jedným kanálom. Pričom všetky ostatné atribúty zostávajú zachované.

Uzol *CountTotal* len vypočítava celkovú konverziu za dané časové obdobie bez aplikovania atribučného modelu. Je to z dôvodu, že pri zobrazovaní štatistík je vhodné porovnanie množstva konverzie po aplikovaní modelu a celkového množstva konverzie. Ide o pomocný údaj pre vizualizáciu. Táto hodnota je potom v uzle *TotalAndCloned* pripojená ku každému



Obrázek 7.9: Transformácie pre metriku s aplikovaným modelom

riadku v toku dát spolu s hodnotou pozície, na ktorej sa kanál nachádzal v rámci konverznej cesty.

Funkčné bloky nazvané *JoinDim1* a *JoinDim2* sú zhodné s krokom v predchádzajúcej transformácii a dochádza v ňom k pripojeniu dát dimenzií k práve spracováanej metrike.

Ďalšou časťou je načítanie modelov pomocou dvoch uzlov - *GetWeights* a *FormatWeights*. Pripojenie hodnôt k hlavnému toku sa deje pridaním váhy na základe pozície. Každý riadok reprezentuje kanál a jeho pozíciu v konverznej ceste, a tak je na základe zhodnosti pozície kanála v konverznej ceste s pozíciou váhy v modeli táto váha priradená k dátam. Tým sa získa tok, ktorý obsahuje dáta pre počítanie metriky napr. počet konverzií, dáta dimenzií a nakoniec identifikáciu kanála a jeho váhu.

V uzle *CountMetric* dôjde k zoskupeniu dát a využije sa agregáčna funkcia suma, ktorá pre danú skupinu získa celkovú hodnotu konverzie. Odhliadnuc od dimenzií sa v tomto uzle vypočíta celková konverzia pre konkrétny kanál na konkrétnej pozícii. V nasledujúcom kroku dôjde k výpočtu množstva konverzie po aplikovaní váhy a v ďalšom kroku sa spočíta celková konverzia, avšak pri zoskupovaní sa už neuvažuje o pozícii kanála v konverznej ceste

7.2.3 Problémy pri vývoji

Počas vývoja projektu došlo k problémom, ktoré súvisia s použitím platformy. Jedným je problém opätovného použitia častí transformácie, druhým bol problém s funkčným blokom, ktorý spracúva dáta vo formáte JSON.

- Z predchádzajúcich uvedených transformácií je vidieť, že podgrafy transformácie pre filtre, dimenzie a konfiguráciu sú úplne rovnaké. Problémom je, že niektoré podtransformácie nemôžu byť jednoducho konfigurované a je ich nutné klonovať a upravovať ručne pri každom použití.
- Problém s funkčným blokom pre transformáciu JSON na tok dát spočíval v tom, že umožňoval prijať a pracovať s veľmi malým vstupom. Rádovo to bolo asi 50 záznamov v tabuľke pre návštevy. Tento problém bol vyriešený použitím funkčného bloku pre vykonávanie programu v jazyku Javascript. V ňom bola implementovaná funkcia, ktorá spracuje odpoveď zo servera a transformuje ju na výstupný tok.

7.3 Implementácia vytvárania štatistík

Pre vytváranie štatistík bol použitý už spomínaný nástroj platformy Pentaho s názvom Pentaho report designer. Ide o desktopovú aplikáciu, ktorá umožňuje prijať dáta z rôznych zdrojov a zobraziť ich definovaným spôsobom.

Pri vytváraní štatistík sa postupovalo podľa dokumentácie a daného postupu práce, ktorý odporúča. Ten je zhrnutý do nasledujúcich krokov.

- Vytvorenie zdroja dát - v tomto prípade je zdrojom dát transformácia vytvorená v Pentaho data integration. Pri vytváraní zdroja sa špecifikuje súbor transformácie, uzol v transformácii, v ktorom sa dáta získajú a nakoniec parametre transformácie.
- Umiestnenie elementov do súhrnu. Tento krok zahŕňa umiestňovanie elementov štatistiky napr. dátové pole, grafy, textové polia, obrázky a ďalšie.
- Aplikovanie formátovania na elementy. Spojenie podštatistík do jednej hlavnej štatistiky.
- Definovanie funkcií a formúl pre úpravu hodnôt dátových elementov.

Aký spôsobom sú jednotlivé kroky vytvárané, je možné nájsť v dokumentácii pre platformu. Preto text pre získanie viacerých detailov o vytváraní súhrnov odkazuje na dokument [\[2\]](#).

Kapitola 8

Testovanie a výstupy

Po predstavení implementačných detailov sú v nasledujúcej časti textu predstavené výstupy a ich testovanie. Pôvodne malo k testovaniu dôjsť na produkčných serveroch, ale technické problémy pri príprave testovacieho prostredia spôsobili, že nástroj bol testovaný len lokálne pomocou ručne generovaných dát.

8.1 Základné metriky

V prvej fáze vytvárania projektu bola vytvorená transformácia, ktorá slúži na zobrazovanie základných metrík, ktorými sú návštevy a zobrazenia stránok. Ako bolo spomenuté, je na tieto metriky možné aplikovať filter a dimenzie. Pre vizualizovanie výstupov bol použitý nástroj Pentaho report designer.

Výstupy týchto základných metrík sú zobrazené na obrázku [8.1](#).

Ten predstavuje pohľad, v ktorom sú použité metriky s dimenziami. V grafe návštev je to metrika návštev s dimenziou dátumu. V časti najobľúbenejší obsah ide o metriku počtu zobrazení stránok s dimenziou stránky. Pri tejto metrike je dimenzia načítavaná z tabuľky dimenzií, zatiaľ čo v prípade návštev je dimenzia získaná z tabuľky faktu.

8.2 Konverzia s modelom

Pre demonštráciu správnej analýzy dát je nutné v krátkosti popísať dátovú sadu. Celkový počet konverzií za sledované obdobie bolo sedem. Z toho bola dvakrát konverzná cesta dlhšia ako jedna. Jedna z týchto ciest v sebe zahŕňa kanály search a referral, druhá cesta je zložená z kanálov email, search a email. Pri oboch cestách vznikla práve jedna konverzia. Ďalšie kanály generovali pri konverznej ceste o dĺžke jedna nasledujúce množstvo konverzie:

- referral - 1
- search - 1
- referral - 2
- direct - 1

Pre zobrazovanie výsledkov konverzie boli vytvorené tri pohľady. Prvý umožňuje zobraziť, ktoré kanále generujú v danom časovom horizonte najviac konverzie po aplikovaní



Obrázek 8.1: Základné metriky

špecifického modelu. V pohľade je možné vidieť množstvo konverzie pre daný kanál po aplikácii modelu a pomer medzi konverziou po aplikácii modelu a celkovou konverziou. Pri tomto pohľade je možné nastaviť pomocou parametra pohľadu akékoľvek rozdelenie konverzie.






V časti 4.2.1 textu sa nachádzajú popísané modely, niektoré použijeme pre demonštráciu výsledkov testovania. Prvým je model poslednej interakcie. Výstup je možné vidieť na obrázku 8.2.

Channel	Conversion with first model	Conversion/Total	Percentage
search	2		28%
referral	2		28%
referral	1		14%
email	1		14%
direct	1		14%

Obrázek 8.2: Konverzia po aplikácii modelu poslednej interakcie

Po zvolení modelu, ktorý prisudzuje zásluhy jednému kanálu sa vyberie model s prisúdením zisku viacerým kanálom, napríklad lineárny model. Pri ňom boli pomery nastavené tak, že každý z kanálov dostal rovnakú časť. Keďže najdlhšia konverzná cesta mala dĺžku

3, tak sa zásluhy sa rozdelia rovnomerne medzi tri kanále, t.j. 33% : 33% : 33%. Výstup tohto modelu je na obrázku 8.3.

Channel	Conversion with first model	Conversion/Total	Percentage
search	0.99		14%
referral	0.99		14%
email	0.66		9%
referral	0.33		4%
direct	0.33		4%

Obrázek 8.3: Konverzia po aplikácii modelu lineárnej interakcie

V ďalšom pohľade je možné zobraziť dva modely vedľa seba a porovnať ich konverzie. Na obrázku 8.4 je možné vidieť porovnanie modelu poslednej a prvej interakcie. Keďže sa

Channel	Conversion with first model	Conversion with second model
search	2	0
referral	1	0
referral	2	0
email	1	1
direct	1	0

Obrázek 8.4: Konverzia pre dva modely

opäť uvažujú tri kanále v rámci modelu a keďže len jedna konverzná cesta mala dĺžku tri tak je pri modeli prvej interakcie vidieť, ktorý kanál získal v tej chvíli zásluhy.

8.3 Konverzia bez modelu

Posledný pohľad zobrazuje množstvo konverzií rozdelených podľa počtu dní, ktoré uplynuli odkedy návštevník prišiel na stránku, až po vykonanie konverzie. Opäť ide o jednoduché tabuľkové zobrazenie, tak ako je to možné vidieť na obrázku 8.5.

Time lag

Days to first visit	Conversions	Conversions value
0	2	20
1	1	10
2	2	20
3	1	10
4	1	10

Obrázek 8.5: Konverzia rozdelená podľa počtu dní od prvej návštevy po konverziu

Kapitola 9

Záver

Cieľom projektu bolo vytvorenie nástroja, ktorý má slúžiť predovšetkým na určovanie atribúcie marketingových aktivít a tým pomáhať pri rozhodovaní, ktoré marketingové aktivity sú skutočne efektívne na ceste ku konverzii. Nástroj by sa mal okrem určovania atribúcie stať plnohodnotným analyzátorom dát, ktoré sa vyskytujú pri prevádzke webovej aplikácie. Niektoré metriky sa podarilo počas vývoja implementovať, ale tie boli vytvárané v rámci konceptu, ktorý potvrdzoval, že jednotlivé komponenty nástroja pôjdu skutočne spojiť.

9.1 Zhodnotenie projektu

Oblasť webovej analytiky je veľmi široká a projekt pokrýva len jej malú časť. Dá sa povedať, že vytvorený nástroj je odrazovým mostíkom pre budúci projekt.

Slúži ako koncept pre potvrdenie, že je možné jednotlivé komponenty, ktoré boli použité, prepojiť a získavať požadované informácie z vytvoreného nástroja.

Projekt bol zadaný firmou Anup s.r.o., ktorú zastupoval Ing. Štěpán Klokočka. Firma mala pripraviť testovacie prostredie pre nástroj, avšak toto nebolo splnené a týmto projekt nebolo možné sledovať postupný vývoj a korigovať jeho smerovanie.

9.2 Aktuálny stav

Projekt sa nachádza v stave, kedy boli splnené špecifikované podmienky. Nástroj dokáže určovať atribúcie marketingových kampaní. Okrem toho umožňuje meranie základných metrík o návštevnosti webu akými sú:

- zobrazenia stránok,
- návštevy,
- návštevníci,
- informácie o obchodných transakciách,
- informácie o zdrojoch návštevnosti.

Dáta sú ukladajú pomocou nástroja ElasticSearch. Použitá platforma Pentaho umožňuje dáta analyzovať a vhodným spôsobom ich zobrazovať.

Pre spracovanie a analýzu dát sú použité transformácie, ktoré sú súčasťou nástroja Pentaho data integration. Transformácie zastupujú ETL proces, počas ktorého dochádza

k získaniu dát zo zdroja, transformácii dát pre ďalšie použitie a postúpenie dát cieľu, ktorý s ním ďalej pracuje. V tomto prípade je cieľom Pentaho report designer, ktorý získané a spracované dáta zobrazuje.

Pre demonštráciu boli v tomto nástroji vytvorené jednoduché pohľady na dáta. Avšak tieto pohľady môžu byť editované, alebo je pomocou spomenutého nástroja možné vytvárať nové zobrazenia dát.

Aktuálne umožňuje projekt prezerať výsledky uložené lokálne pomocou desktopovej aplikácie.

9.3 Budúcnosť projektu

Projekt má miesta, v ktorých by mohlo dôjsť k ďalšiemu vývoji a vylepšovaniu. V prvom rade by mal byť dostatočne otestovaný a odladený pre použitie na produkčných serveroch, zahŕňajúc testovanie pri záťaži.

Ako bolo spomenuté v predchádzajúcej časti, projekt je konceptom, ktorý potvrdzuje prepojenie použitých komponentov.

V oblasti webovej analytiky by sa mohli doplniť metriky, pre sledovanie napríklad pohybov myši na stránkach alebo kliknutia na elementy stránky pre lepšie pochopenie práce návštevníka s danou stránkou. Pri ďalšom vývoji by mohli byť pridané metriky lojality návštevníkov alebo informácie o softvéri, ktorý návštevníci používajú pri práci s webovou aplikáciou. Posledné spomenuté dáta sú skôr vecou vývoja vizualizácie, pretože nástroj tieto dáta umožňuje merať, ale neumožňuje ich zobrazíť, keďže nešlo o primárny cieľ projektu.

V budúcnosti by malo byť možné prezerať štatistiky, ktoré sú uložené na serveri a spravované centrálné, prípadne umožniť viacerým ľuďom podieľať sa na vytváraní jednej štatistiky.

Literatura

- [1] *What is elasticsearch?* [online], 2013, poslední revize 11.5.2013 [cit. 2013-05-11].
URL <http://www.elasticsearch.org/overview/>
- [2] *Pentaho Report Designer User Guide*. c2013, poslední revize 26.4.2013 [cit. 11.5.2013].
URL http://infocenter.pentaho.com/help/topic/report_designer_user_guide/report_designer_user_guide.pdf
- [3] Creese, G.; Veytsel, A.: *Web Analytics: Making Business Sense of Online Behavior*. The Aberdeen Group, Inc., 2002.
- [4] Google: *Jak program AdWords funguje*. [online], 2013 [cit. 2013-05-01].
URL <http://support.google.com/adwords/bin/answer.py?hl=cs&answer=1752122>
- [5] Hewstone, M.; Fincham, F. D.; Jaspars, J.: *Attribution theory and research: Conceptual developmental and social dimensions*. Academic Press, 1983, ISBN 0123809800.
- [6] Holzner, S.: *Physics for dummies*. Wiley, John & Sons, Incorporated, 2005, ISBN 0470903244.
- [7] Kaushik, A.: *Web Analytics: An hour a day*. Wiley Publishing, Inc., 2007, ISBN 0470130652.
- [8] Lohr, S.: For impatient web users, an eye blink is just too long to wait. *The New York Times*, 29 Únor 2012.
URL http://www.nytimes.com/2012/03/01/technology/impatient-web-users-flee-slow-loading-sites.html?pagewanted=all&_r=0
- [9] Nielsen, J.: *Usability engineering*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1993, ISBN 0125184050.
- [10] Open web analytics: *Comparsion*. [online], 2013, [cit.2013-01-05].
URL http://www.openwebanalytics.com/?page_id=158
- [11] Peterson, E. T.: *Web analytics demystified*. Celio Group Media and CafePress, ISBN 0-9743584-2-8.
- [12] RichMedia: *CACHE-BUSTING*. [online], 2006, poslední revize 8.9.2006 [cit. 2013-05-01].
URL <http://www.247realmedia.com/EN-US/tech/oas/RichMedia/local/technotes/cache.html>

- [13] Seznam.cz: *Kde se Sklik zobrazuje*. [online], 2013, [cit.2013-05-01].
URL <http://napoveda.sklik.cz/cz/zaciname-inzerovat/kde-se-sklik-zobrazuje/>
- [14] Seznam.cz: *Radenie inzerátov*. [online], 2013, [cit.2013-05-01].
URL <http://napoveda.sklik.cz/cz/zaciname-inzerovat/inzeraty-razeni/>
- [15] Sherman, C.: The State of Search Engine Marketing 2006. *Search Engine Land*, 8 Únor 2007, [cit.2013-01-05].
URL <http://searchengineland.com/the-state-of-search-engine-marketing-2006-10474>
- [16] Slingshot SEO: Valuing digital marketing channels with attribution models. [online], 2011, [cit. 2013-05-01].
URL <http://www.slingshotseo.com/resources/white-papers/valuing-conversions-through-multi-touch-attribution/>
- [17] Warren, J.: ROI in online marketing and web analytics. [PDF], 2011.
- [18] Web Analytics Association: *Web analytics definitions*, 22 Zář 2008.
- [19] webProgress.cz: *Internetový marketing a jeho formy*. [online], 2013, [cit. 2013-05-01].
URL <http://www.vicezakazniku.cz/internetovy-marketing-a-jeho-formy>
- [20] Zakas, N. C.: *How many users have JavaScript disabled?* [online], 2010, poslední revize 13.10.2010 [cit. 2013-05-01].
URL <http://developer.yahoo.com/blogs/ydn/posts/2010/10/how-many-users-have-javascript-disabled/>
- [21] Zákaznický informační servis: *Základní nástroje on-line marketingu*. [online], 2012, [cit. 2013-01-03].
URL <http://zis.antee.cz/index.php?nid=7515&lid=cs&oid=2400888>

Příloha A

Obsah DVD

DVD obsahuje:

- Nástroje Pentaho platformy
 - Pentaho data integration - nástroj pre vytváranie a úpravu transformácií
 - Pentaho report designer - nástroj pre vytváranie súhrnov
- ElasticSearch 0.20.5 - databáza pre ukladanie dát
- textová časť diplomovej práce – zdrojové súbory dokumentov napísane v L^AT_EXu
- transformácie pre získavanie dát,
- štatistiky vytvorené v Pentaho report designer.

Příloha B

Inštalácia

Pre úspešné nainštalovanie prostredia sa odkazujem na dokumentáciu jednotlivých nástrojov:

- *ElasticSearch* - je na DVD umiestnený v priečinku ElasticSearch ako archív. Ten je nutné rozbaľiť na akékoľvek miesto v systéme, kde má užívateľ prístup k zápisu. Potom treba prejsť do tohto priečinka a nasledovným príkazom spustiť databázu.

```
bin/elasticsearch
```

Pokiaľ chcete spúšťať databázu ako službu, použite nasledujúci odkaz pre viac detailov:

```
http://www.elasticsearch.org/guide/reference/setup/installation/
```

- *Pentaho data integration* - je na DVD umiestnený v priečinku Data integration. Archív je nutné rozbaľiť a nie je nutné robiť žiadnu dodatočnú konfiguráciu pre tento nástroj. Po rozbalení nástroja sa jeho dokumentácia nachádza v priečinku *docs/English*. Je tam možné nájsť všetky informácie, ako nástroj ovládať.
- *Pentaho report designer* - je na DVD umiestnený v priečinku Report designer. Pre úspešné fungovanie transformácií z predchádzajúceho nástroja je nutné urobiť nasledujúce kroky:
 - rozbaľiť archív v priečinku Report designer.
 - prekopírovať všetky JAR balíčky a podpriečinky z predchádzajúceho nástroja, ktoré sa nachádzajú v priečinku *libext* okrem podpriečinka s názvom *JDBC*, do priečinka v Report designeri s názvom *lib*. Toto je dôležitý krok. Jeho vynechanie spôsobí, že budú štatistiky hlásiť chyby.
- *Open web analytics* - je na DVD umiestnené v priečinku OWA. Open web analytics je nutné rozbaľiť do priečinka, kde je prístupný webovému serveru kvôli registrovaniu stránok a cieľov. Tento archív obsahuje už vložený databázový ovládač pre komunikáciu s ElasticSearchom. Pre viac detailov ako správne nainštalovať Open web analytics viď.:

```
http://wiki.openwebanalytics.com/index.php?title=Main\_Page
```

Nastavenie databázového spojenie je odlišné od MySQL a je ho nutné vykonať v nasledujúcich krokoch:

- premenovať súbor owa-config-dist.php na owa-config.php,
- nastaviť nasledujúce parametre podľa seba:

- * názov ovládača
`define('OWA_DB_TYPE', 'elastica');`
- * názov databázy v ElasticSearch podľa toho, ako ste ho nakonfigurovali v prvom kroku.
`define('OWA_DB_NAME', 'owa');`
- * názov domény, na ktorej ElasticSearch beží
`define('OWA_DB_HOST', 'localhost');`
- * číslo portu inštalácie ElasticSearch, na ktorom prijíma požiadavky
`define('OWA_DB_PORT', '9200');`

- *Vlastné riešenie* - je na DVD umiestnené v priečinku AttributionTool. Obsahuje štatistiky (súbory s koncovkou prpt) a transformácie, ktoré boli vytvorené pri vývoji nástroja. Tieto súbory štatistík je možné otvoriť v Report designeri a pomocou náhľadu zobraziť grafickú reprezentáciu dát.
- *Nastavenie spojenia pre transformácie* - po skopírovaní priečinku s riešením je v priečinku *reports/metrics* súbor database.properties. V ňom je nutné doplniť informácie pre pripojenie k databáze, napríklad takto:

```
databaseName=owa  
databaseHost=http\://localhost:9200/
```

Opäť je nutné nastaviť informácie podľa konkrétnej databázy.